2017-04-26

Bayesian Models for the Analyzes of Noisy Responses From Small Areas: An Application to Poverty Estimation

Binod Manandhar
Worcester Polytechnic Institute

Bayesian Models for the Analyzes of Noisy Responses From Small Areas: An Application to Poverty Estimation

by

Binod Manandhar

A Thesis

Submitted to the Faculty

of

WORCESTER POLYTECHNIC INSTITUTE

In partial fulfillment of the requirements for the

Degree of Doctor of Philosophy

in

Mathematical Sciences

April 10, 2017

APPROVED:

Professor Balgobin Nandram, Advisor
Department of Mathematical Sciences
Worcester Polytechnic Institute

Jian Zou
Department of Mathematical Sciences
Worcester Polytechnic Institute

Huong N. Higgins
Foisie Business School
Worcester Polytechnic Institute

Professor Dominique Haughton
Department of Mathematical Sciences
Bentley University

Jai Won Choi
Statistical Consultant, Meho Inc.
9504 Mary Knoll Dr., Rockville MD 20850

Contents

List of Figures

List of Tables

Acknowledgments

I would like to express my sincere gratitude to my advisor, Professor Balgobin Nandram, for his continuous support and guidance, and for his motivation throughout my study. His guidance helped me all the time of research and writing of my dissertation. I admire his desire and willing to explore new areas and continuously working hard. It has been an honor to be his student.

I am also very grateful to all of my dissertation committee members for the time and effort they spent in reading and in evaluating my dissertation. Their valuable comments and suggestions have helped to improve dissertation. I am grateful to the Department of Mathematical Sciences for providing financial support for my entire period at WPI, all its faculty and staff for their help and kindness.

I would like to acknowledge professors at University of North Florida (UNF): Pali Sen, and James U. Gleaton; and at Utah State University (USU): Daniel Coster, David Bregenzer, and Justin Heavilin who helped me when it was needed.

I would like to extend gratitude to my mother Ram Kumari Manandhar and my late father Bhim Das Manandhar. My parents always want all their sons get higher education and contribute something to mankind to the best of their abilities. I would like to thank all my brothers and the rest of my family members for their moral and emotional support. I am grateful to all my colleagues and friends in the Nepalmandal as well as in the United States for their support.

This dissertation is dedicated to

LunMadhi Ajima *(Bhadrakali)*

and

awakened souls.

List of Acronyms

BHF	Battese, Harter, and Fuller
CBS	Central Bureau of Statistics
CDF	Cumulative Distribution Function
CPO	Conditional Predictive Ordinates
CPS	Continuous and Positively Skewed
EB	Empirical Bayesian
ELL	Elbers, Lanjouw, and Lanjouw
GB2	Generalized Beta Distribution of second kind
GGamma	Generalized Gamma
HB	Hierarchical Bayesian
LPML	Logarithmic of the Pseudo Marginal Likelihood
LSMS	Living Standards Measurement Study
MCMC	Markov chain Monte Carlo
MH	Metropolis Hastings
MLE	Maximum Likelihood Estimate
NER	Nested Error Regression
NLSS	Nepal Living Standard Survey
PSU	Primary Sampling Unit
SA	Small Area
SAE	Small Area Estimation
SE	Standard Error
VDC	Village Development Committee
WB	World Bank

Abstract

We implement techniques of small area estimation (SAE) to study consumption, a welfare indicator, which is used to assess poverty in the 2003-2004 Nepal Living Standards Survey (NLSS-II) and the 2001 census. NLSS-II has detailed information of consumption, but it can give estimates only at stratum level or higher. While population variables are available for all households in the census, they do not include the information on consumption; the survey has the 'population' variables nonetheless. We combine these two sets of data to provide estimates of poverty indicators (incidence, gap and severity) for small areas (wards, village development committees and districts).

Consumption is the aggregate of all food and all non-food items consumed. In the welfare survey the responders are asked to recall all information about consumptions throughout the reference year. Therefore, such data are likely to be noisy, possibly due to response errors or recalling errors. The consumption variable is continuous and positively skewed, so a statistician might use a logarithmic transformation, which can reduce skewness and help meet the normality assumption required for model building. However, it could be problematic since back transformation may produce inaccurate estimates and there are difficulties in interpretations.

Without using the logarithmic transformation, we develop hierarchical Bayesian models to link the survey to the census. In our models for consumption, we incorporate the 'population' variables as covariates. First, we assume that consumption is noiseless, and it is modeled using three scenarios: the exponential distribution, the gamma distribution and the generalized gamma distribution. Second, we assume that consumption is noisy, and we fit the generalized beta distribution of the second kind (GB2) to consumption. We consider three more scenarios of GB2: a mixture of exponential and gamma distributions, a mixture of two gamma distributions, and a mixture of two generalized gamma distributions. We note that there are difficulties in fitting the models for noisy responses because these models have non-identifiable parameters. For each scenario, after fitting two hierarchical Bayesian models (with and without area effects), we show how to select the most plausible model and we perform a Bayesian data analysis on Nepal's poverty data.

We show how to predict the poverty indicators for all wards, village development committees and districts of Nepal (a big data problem) by combining the survey data with the census. This is a computationally intensive problem because Nepal has about four million households with about four thousand households in the survey and there is no record linkage between households in the survey and the census. Finally, we perform empirical studies to assess the quality of our survey-census procedure.

Chapter 1

Introduction

In this dissertation, we study consumption in the survey, which is used to infer about poverty in a census. Continuous and positively skewed (CPS) data, such as consumption, income, insurance, and loss in numerous applications, are examples of size data. Such data are generally heavy-tailed and skewed to the right. The logarithmic transformation is the most widely used tool to meet the normality assumption for a CPS size data. Once the normality assumption is satisfied, it makes model-building, computation, and further analysis easier. However, the logarithmic transformation for model building could be problematic, so we develop the hierarchical Bayesian models for CPS data without logarithmic transformation. We assume that the variable under study is noiseless (observed without recalling errors) or noisy (observed with recalling errors). We fit the generalized gamma distribution for noiseless data and fit the generalized beta distribution of the second kind (GB2) distribution for noisy data.

There are numerous statistical models and tests which have been developed under the normality assumption of the variable under study. If the variable under study is not normal or approximately normal, then it is standard practice to meet the normality assumption of the variable under study by a transformation. If the response variable y^* is not approximately normal, then it is transformed to meet the normality assumption. The choice of transformation depends upon the nature of distribution of the responses. Here, we give some of the most popular examples of transformations. If the responses are CPS, then the transformation could be the logarithm $y = \ln(y^*),\ y > 0$ or a reciprocal $y = \frac{1}{y^*},\ y \neq 0$. If the responses are negatively skewed, then the transformation could be square $y = (y^*)^2$

or exponential $y = e^{y^*}$. If the responses are counts, then the transformation could be the square root $y = \sqrt{y^*},\ y \geq 0$.

If we have used a transformation to build a model, the usual way to get estimates back to their original scale is to perform back-transformation. Does back-transformation give a correct distribution of the response variable? Furthermore, what if the normality assumption fails? Feng et al. (2013, 2014) discussed the problems with using the logarithmic transformation for positively skewed data. Since the logarithmic transformation could be problematic, it would be better if we had a model that could give better estimates without the logarithmic transformation of the variable under study. Our purpose is to develop a statistical model without logarithmic transformation for CPS data.

We build hierarchical Bayesian models for the CPS data without logarithmic transformation and predict the responses for both sampled and non-sampled units. Prediction arises in many problems of statistical analysis, with one of them being small area estimation (SAE). This dissertation focuses on giving estimates for small areas. SAE is essential for different sectors like government agencies, development partners, planners, and researchers for many purposes like developmental planning. A small area (SA) could be a contiguous or non-contiguous geographical area or a class of characteristics for which we desire estimates. SAE refers to a collection of statistical techniques designed for improving sample survey estimates through the use of auxiliary information (Rao and Molina, 2015). In SAE a statistical model is fitted to the survey data that are enumerated around the same time as the census. This model is used to predict a variable not collected in the census, based on the variables that are collected in both survey and census (Haslett et al. 2006, 2012). A survey or experiment may have a large or small sample size according to the budget, time, and detailed information needed. In general, surveys with smaller sample sizes tend to collect more detailed information than surveys with larger sample sizes. Surveys with smaller sample sizes are mostly designed to study a specific question. For example, a national income survey is designed to study income; an agriculture survey is designed to study agricultural production; and a living-standards survey is designed to study well-being. These surveys generally have fewer samples, but collect detailed information on a specific topic. Because of the smaller sample sizes, these surveys cannot give estimates for small areas. Similarly, very large-scale surveys like population censuses, agricultural censuses, industrial censuses, or any other large surveys are not designed to give estimates like income, health, or welfare. Large-scale surveys tend to collect only general information like age, gender, literacy, birth, or death. Hence a large-scale survey could give SAE but only for the limited general information it contains. The good thing for statisticians is that small-sample surveys also collect the same general information as large-sample surveys. Examples of these common types of general information are demographic statistics such as age, gender, and household size; education statistics such as highest education level; welfare statistics such as type of house, type of cooking fuels, and availability of facilities such as phone, TV, and internet. In a small-sample survey, the general information is collected for the completeness of the survey or to check the consistency of the data. If we have information with the same covariates in two or more surveys, which come from the same population and have the same distribution, then we can exploit it for SAE. These common covariates play the important role of a bridge in SAE. To facilitate the prediction, we introduce covariates in our model to explain the response variable.

The goal of this dissertation is to fit models without the logarithmic transformation to the CPS data from small areas, which is accommodated by using the positively skewed density function with support $(0, \infty)$ for modeling, and introduce covariates in the model. We fit models with the exponential, the gamma, the generalized gamma (GGamma), and the GB2 distribution. GB2 is the mixture of two generalized gamma distributions, where the distribution of the response variable is mixed with the distribution of its rate parameter, which has another generalized gamma distribution. Therefore, GB2 has one more fold of distribution than the generalized gamma distribution, and the true rate parameter of the response variable is hidden.

In this dissertation we consider two possibilities for the available response data, with or without noises. This noise could have been introduced into the response data as random errors or recalling errors. We choose suitable distributions for the model according to whether the responses are noisy or not. We have chosen the exponential, the gamma, and the generalized gamma distributions for noiseless data and the GB2 distribution for noisy response data. We fit three special cases of the GB2 distribution for the noisy responses: the mixture of exponential and gamma distributions, the mixture of two gamma distributions, and the mixture of two generalized gamma distributions. We fit the same model twice, first the model without random area effects and second the model with random area effects, regardless of whether the responses are noisy or not.

The proposed models in this dissertation have posterior densities and conditional posterior densities in complex form, which makes parameter drawing tedious. To ease these difficulties, we have used the second-order Taylor's series approximation that will help us by providing approximated multivariate normal distribution for a large set of parameters.

Model adequacy has been checked by Bayesian cross-validation as proposed by Gelfand, Dey, and Chang (1992), which involves prediction of y_i of the data $\boldsymbol{y}$, when the components $\boldsymbol{y}_{(i)}$ are used. The vector $\boldsymbol{y}_{(i)}$ denotes all observations of $\boldsymbol{y}$ except the i^{th} observation, y_i. The cross-validation approach finds $p(y_i|\boldsymbol{y}_{(i)})$, called conditional predictive ordinates (CPO), defined by Box (1980). We have used summary statistics of the CPOs, the logarithm of the pseudo-marginal likelihood (LPML) to compare models. We apply our models to the welfare consumption CPS data from NLSS-II, 2003–2004, by assuming responses are noiseless and then assuming responses are noisy. From the fitted models we choose the best two models, one best model by assuming noiseless responses and another best model by assuming noisy responses. The best selected models are then applied to the census data for the prediction of responses and provide the SAE. By showing which models fits better, we are actually showing whether noisy or noiseless model fits better.

1.1 Literature Review

In this section, we will review some of the existing literature on the predictive models for the log-transformed CPS data for small area estimation both in Bayesian and non-Bayesian paradigms.

1.1.1 Non-Bayesian Paradigm (World Bank, ELL Model)

The World Bank method, popularly known as the Elbers, Lanjouw, and Lanjouw (ELL) method (2001, 2003), has been commonly used in SAE of welfare measures based on the estimation of consumption or expenditure. It had already been applied in 60 countries by 2011 (The World Bank, 2013). The model response variable is a log-transformed positively skewed welfare variable. The ELL model was designed specifically for the SAE of poverty measures based on per capita consumption or expenditure. Consumption surveys such as living-standards surveys are specific surveys and designed to collect information on the welfare of citizens of a country. Although the sample size could be small, the survey is extensive and covers different sections like food consumption, non-food consumption, expenditures, income levels and sources, health, and demography. But, survey sample size is typically very small and designed to give estimates for stratum or larger areas. Those direct estimates for larger areas are readily available from the survey data, but because of the small sample size, it cannot give estimates in small areas. It is obvious that most of the small areas have no sample units selected in the survey.

Following the SAE technique, ELL uses information from surveys such as the living standards survey and large surveys such as the population census. The ELL model has been used for SAE in numerous countries since 2001. Some of the countries where ELL model SAE has been implemented are Brazil, 2001; South Africa, 2002; Thailand, 2003; Bangladesh, 2003; Cambodia, 2004 and 2012; the Philippines, 2005; and Nepal, 2006, 2013 (Haslett et al., 2012).

The ELL model is a nested error regression model, which is built using survey data, and provides measures of SAE, especially for small political boundaries. The survey data, from which the model is built, has information on both responses and covariates. The response variable is the log-transformed positively skewed per capita consumption or welfare variable. A nested error multivariate linear regression model between the response variable Y and covariates X is

$$\boldsymbol{Y} = X'\boldsymbol{\beta} + \boldsymbol{\zeta},$$

where $\boldsymbol{\beta}$ is the regression coefficient and $\boldsymbol{\zeta}$ is a random error that cannot be explained by the covariates. Assume that there are n observations sampled from ℓ small areas and y_{ij} denotes the response for the j^{th} unit in the i^{th} area with the corresponding covariates $\boldsymbol{x_{ij}}$, $i = 1, \cdots, \ell,\ j = 1, \cdots, n_i$. The total random error ζ_{ij} associated with the response y_{ij}, can be decomposed into the sum of two errors, the area effect u_i of the i^{th} area and the unit effect e_{ij} as

$$\zeta_{ij} = u_i + e_{ij}.$$

If we assume that this relationship between the response variable $\boldsymbol{Y}$ and covariates X holds in non-sampled units, then we can apply this model in non-sampled units to predict responses. Then, the above regression model can be written more explicitly as a nested error model (Battese, Harter, and Fuller, 1988)

$$y_{ij} = \boldsymbol{x'_{ij}}\boldsymbol{\beta} + u_i + e_{ij}, \quad i = 1, \cdots, \ell,\ j = 1, \cdots, n_i,$$

where u_i is the error term held in common by the units of the i^{th} area, and e_{ij} is the unit error of the j^{th} unit in the i^{th} area. The area-level error has variance σ_u^2 and the unit-level error has variance σ_e^2. The ELL model assumes that the unit error e_{ij} is heteroscedastic and the area effects errors u_i come from the same distribution. A heteroscedastic variable is modeled by allowing the function of the variance σ_e^2 to be linearly related through the covariates $\boldsymbol{Z}$ with a regression coefficient α and a random error $\boldsymbol{r}$

$$g(\sigma_e^2) = \boldsymbol{Z'}\boldsymbol{\alpha} + \boldsymbol{r},$$

where g(.) is a link function. Fujii (2004) used a more general logistic-type link function to model the heteroscedastic variable, which is given by

$$\ln\left(\frac{\hat{e}_{ij}^2}{A - \hat{e}_{ij}^2}\right) = \boldsymbol{z'_{ij}}\boldsymbol{\alpha} + r_{ij},$$

where $A = 1.05 \times \max_{ij}\{\hat{e}_{ij}^2\}$.

The fitted variances $\hat{\sigma}^2_{e,ij}$ are estimated from the above model. This estimated variance of household error is then used to standardize the unit residual $\hat{e}^*_{ij} = \hat{e}_{ij}/\hat{\sigma}_{e,ij}$.

1.1.2 Empirical Bayesian Approach

The empirical Bayesian (EB) model for SAE is a nested error model which decomposes the total error into the sum of the area error and the unit error. The EB model for small area estimation uses the logarithmic transformed welfare variable as the response variable with covariates. We review briefly an EB model as explained in Molina and Rao's (2010) paper.

Let there be L small areas and assume that each small area has N_i enumeration units, let Y_{ij} be the response variable, and $\boldsymbol{x_{ij}}$ is a vector of p covariates in the i^{th} area and j^{th} unit. The area effect is u_i, the unit residual is e_{ij}, and they are independent. The model is

$$y_{ij} = \boldsymbol{x'_{ij}\beta} + u_i + e_{ij}, \quad i = 1, \cdots, \mathrm{L}, \quad j = 1, \cdots, N_i.$$

The informative priors for area effect and unit residual are

$$u_i|\sigma_u^2 \stackrel{\text{iid}}{\sim} N(0, \sigma_u^2), \quad e_{ij}|\sigma^2 \stackrel{\text{iid}}{\sim} N(0, \sigma_e^2).$$

Then the vectors $\boldsymbol{y_i}, i = 1, \cdots, L$ are independent with

$$\boldsymbol{y_i} \stackrel{\text{iid}}{\sim} N(\boldsymbol{\mu_i}, V_i),$$

where

$$\boldsymbol{\mu_i} = \boldsymbol{x}'_i\boldsymbol{\beta}, \text{ and } V_i = \sigma_u^2 \mathbf{1}_{N_i}\mathbf{1}'_{N_i} + \sigma_e^2 I_{N_i},$$

$\mathbf{1}_{N_i}$ denotes the column vector of ones of size N_i, and I_{N_i} is the $N_i \times N_i$ identity matrix.

Let the response vector in the population be $\boldsymbol{y} = (Y_1, \cdots, Y_N)'$ and $\boldsymbol{y_s}$ corresponding to the sampled units and $\boldsymbol{y_r}$ corresponding to the non-sampled units. For the i^{th} area, the response vector decomposed into sampled and non-sampled is $\boldsymbol{y}_i = (\boldsymbol{y'_{is}},\ \boldsymbol{y'_{ir}})'$. When the sample size is greater than zero, the corresponding decomposition for covariates, the mean vector and the covariance matrix are respectively X_i, μ_i, and V_i. Then, the distribution of non-sample responses given sampled responses are

$$\boldsymbol{y_{ir}}|\boldsymbol{y_{is}} \sim N(\boldsymbol{\mu_{ir|s}}, V_{ir|s}),$$

where

$$\boldsymbol{\mu_{ir|s}} = \boldsymbol{X_{ij}\beta} + \sigma_u^2 \mathbf{1}_{N_i-n_i}\mathbf{1'}_{n_i} V_{is}^{-1}(\boldsymbol{y_{is}} - \boldsymbol{X_{is}\beta}),$$
$$V_{ir|s} = \sigma_u^2(1-\gamma_i)\mathbf{1}_{N_i-n_i}\mathbf{1'}_{N_i-n_i} + \sigma_e^2 I_{N_i-n_i},$$

for $V_{is} = \sigma_u^2 \mathbf{1_{n_i}}\mathbf{1'_{n_i}} + \sigma_e^2 I_{n_i}$ and $\gamma_i = \frac{\sigma_u^2}{\sigma_u^2+\sigma_e^2/n_d}$. Since $\boldsymbol{y_i}, i = 1, \cdots, L$, are assumed independent, $\boldsymbol{y_{ir}}|\boldsymbol{y_{is}}$ and $\boldsymbol{y_{ir}}|\boldsymbol{y_s}$ have the same distribution. The covariance matrix $V_{ir|s}$ corresponds to the covariance matrix of a vector $\boldsymbol{y}_{ir}$ generated by the model

$$\boldsymbol{y}_{ir} = \boldsymbol{\mu_{ir|s}} + \nu_i \mathbf{1}_{N_i-n_i} + \epsilon_{ij},$$

with new random area effects ν_i and unit residual ϵ_{ij} that are independent and satisfy

$$\nu_i \sim N(0, \sigma_u^2(1-\gamma_i)), \quad i = 1, \cdots, L \quad \text{and} \quad \boldsymbol{\epsilon}_{ir} \sim N(\mathbf{0}_{\boldsymbol{N_i-n_i}}, \sigma_e^2 I_{N_i-n_i}).$$

If any small area i is not sampled, then $Y_{ij}^{(b)}$, for $j = 1, \cdots, N_i$, are generated by bootstrap from

$$Y_{ij} = \boldsymbol{x'_{ij}\beta} + u_i^* + e_{ij}^*,$$

where

$$u_i^* \overset{\text{iid}}{\sim} N(0, \hat{\sigma}_u^2), \quad e_{ij}^* \overset{\text{iid}}{\sim} N(0, \hat{\sigma}_e^2),$$

and u_i^* is independent with e_{ij}^*.

1.1.3 Hierarchical Bayesian Approach

The hierarchical Bayesian model for continuous and positively skewed data is shown in the paper, by Molina, Nandram, and Rao in 2014. This method does not need Markov chain Monte Carlo (MCMC) methods. We review this method in brief as follows.

Let there be L small areas and each small area have N_i enumeration units. The model response is the log transformed positively skewed welfare variable, and there are p covariates. As in the ELL model and the EB model it also fits the nested error regression, BHF model. The total error is decomposed into the sum of two errors, the area error and the unit error.

Let the log transformed positively skewed welfare variable in the population be Y_{ij} with covariate $\boldsymbol{x}_{ij},\ i = 1, \cdots, \mathrm{L},\ j = 1, \cdots, N_i$. The nested error regression model is

$$Y_{ij} = \boldsymbol{x'_{ij}\beta} + u_i + e_{ij}, \quad i = 1, \cdots, L; \quad j = 1, \cdots, N_i,$$

where $\boldsymbol{x_{ij}}$ is the $p \times 1$ vector of covariates for j^{th} unit within the i^{th} area, β is the $p \times 1$ regression coefficients vector, u_i is a random area effect of the i^{th} area, and e_{ij} is the unit model error corresponding to the response y_{ij}. It is assumed that given parameters, the area effects u_i and unit errors e_{ij} are independent as in the empirical Bayesian model. The priors for u_i and e_{ij} are informative and $\boldsymbol{\beta}, \sigma^2, \text{and}\, \rho$ have non-informative priors

$$u_i|\sigma_u^2 \overset{\text{iid}}{\sim} N(0, \sigma_u^2), \quad e_{ij}|\sigma^2 \overset{\text{iid}}{\sim} N(0, \sigma^2 w_{ij}^{-1}), \quad e_{ij}|\sigma^2 \overset{\text{iid}}{\sim} N(0, \sigma^2 w_{ij}^{-1}), \quad \pi(\boldsymbol{\beta}, \sigma^2, \rho) \propto \frac{1}{\sigma^2},$$

where $\rho = \frac{\sigma_u^2}{\sigma_u^2 + \sigma^2}$ is an intra-class correlation coefficient. Then the reparameterized model is

$$\begin{aligned} Y_{ij}|u_i, \boldsymbol{\beta}, \sigma^2 &\overset{\text{ind}}{\sim} N\left(\boldsymbol{x'_{ij}\beta} + u_i,\ \ \sigma^2 w_{ij}^{-1}\right) \\ u_i|\rho, \sigma^2 &\overset{\text{ind}}{\sim} N\left(0, \frac{\rho}{1-\rho}\sigma^2\right). \end{aligned}$$

In total area L, we have sampled ℓ areas and $(L - \ell)$ areas that are not in the sample. In each sampled area n_i units have been selected and $N_i - n_i$ units are not selected. The heteroscedasticity of enumeration level error is denoted by $w_{ij} > 0$. Let s be the set of units selected in the sample and r be the set of units not selected. Without loss of generality, let us assume we have $n_i > 0$ for sampled areas $i = 1, \cdots, \ell$ and $n_i = 0$ for areas $i = \ell{+}1, \cdots, L$ not in the sample.

The posterior distribution is given by

$$\pi(\boldsymbol{u}, \boldsymbol{\beta}, \sigma^2, \rho | \boldsymbol{y_s}) \propto \left[\prod_{i=\ell+1}^{L} \pi(\boldsymbol{u_i} | \boldsymbol{\beta}, \sigma^2, \rho) \right] \left(\frac{1-\rho}{\rho} \right)^{\ell/2} (\sigma^2)^{-\left(\frac{\ell+n}{2}+1\right)}$$
$$\times \ \exp \left(-\frac{1}{2\sigma^2} \sum_{i=1}^{\ell} \left[\sum_{j=1}^{n_i} w_{ij} \left(y_{ij} - \boldsymbol{x'_{ij}} \boldsymbol{\beta} - u_i \right)^2 + \frac{1-\rho}{\rho} u_i^2 \right] \right).$$

It can be represented as the product of conditional probabilities using the multiplication rule of probability as given below, and the samples could be drawn without using Markov chain Monte Carlo (MCMC). Computation in this model is faster as we do not have to use MCMC. The joint posterior density function can be written as

$$\pi(\boldsymbol{u}, \boldsymbol{\beta}, \sigma^2, \rho | \boldsymbol{y_s}) = \pi_1(\boldsymbol{u} | \boldsymbol{\beta}, \sigma^2, \rho, \boldsymbol{y}_s) \ \pi_2(\boldsymbol{\beta} | \sigma^2, \rho, \boldsymbol{y}_s) \ \pi_3(\sigma^2 | \rho, \boldsymbol{y}_s) \ \pi_4(\rho | \boldsymbol{y}_s),$$

where the conditional densities $\pi_1(\boldsymbol{u} | \boldsymbol{\beta}, \sigma^2, \rho, \boldsymbol{y}_s), \pi_2(\boldsymbol{\beta} | \sigma^2, \rho, \boldsymbol{y}_s),$ and $\pi_3(\sigma^2 | \rho, \boldsymbol{y}_s)$ are in simple closed form and samples could be drawn from their respective distributions. The conditional posterior density $\pi_4(\rho | \boldsymbol{y}_s)$ is not in simple form and samples are drawn using the grid method (e.g., see Nandram and Yin 2016a, 2016b).

1.1.4 Other Approaches

The data from two or more than two levels can be modeled by introducing effects from more than one level. The four-level model with the smallest enumeration unit level (first level), $i = 1, \cdots, n_j$, second level, $j = 1, \cdots, m_k$, third level, $k = 1, \cdots, r_\ell$, and fourth level, $\ell = 1, \cdots, 12$ Nguyen, Haughton, Hudson and Boland (2010) and Nguyen (2014) is

$$\begin{aligned} y_{ijk\ell} &= \beta_{0\,jk\ell} + \sum_p \boldsymbol{\beta}_{p\,jk\ell} \, \boldsymbol{x}_{p\,ijk\ell} + \epsilon_{ijk\ell} \\ \beta_{0\,jk\ell} &= \gamma_{00} + \boldsymbol{\gamma}_{01} \, \boldsymbol{z}_{jk\ell} + f_{0\,\ell} + \nu_{0\,k\ell} + u_{0\,jk\ell} \\ \beta_{p\,jk\ell} &= \gamma_{p0} + \boldsymbol{\gamma}_{p1} \, \boldsymbol{z}_{jk\ell} + f_{p\,\ell} + \nu_{p\,k\ell} + u_{p\,jk\ell}, \end{aligned}$$

where, $y_{ijk\ell}$ is the log-transformed response and $x_{p\,ijk\ell}$ is the p^{th} covariate at the first level, the i^{th} observed unit in the j^{th} second level of the k^{th} third level and the ℓ^{th} fourth level. The $\boldsymbol{z}_{jk\ell}$ denote the covariates at the second level j. The $\beta_{0\,jk\ell}$ and $\beta_{p\,jk\ell}$ are the regression intercepts and the regression coefficients. The $\epsilon_{ijk\ell}$ is the random error. The area specific error terms: u_{pjkl} is random error at the second level, ν_{0kl}, and ν_{pkl} are random errors at third level and f_{0l}, and f_{pl} are random error terms at the fourth level. Errors are assumed to have a mean of zero and a constant variance.

Other approaches also exist for fitting asymmetric data. A three-parameter asymmetric Laplace distribution is one way for fitting quantiles, quantile regression for data analysis (Yu and Zhang, 2005). Another approach is the beta regression model. Beta distribution is very flexible over a $(0, 1)$ range and allows for an asymmetric sampling distribution. Beta regression models have been used for studying poverty and inequality related to small area parameters and approximated Bayesian inference relaying on the MCMC algorithm (Fabrizi, Rosaria and Trivisano, 2016). The concise combination of EB and HB nested error regression models together are provided by Rao and Molina (2016), which we have discussed above.

If our interest is in the prediction of proportion for the characteristic of our CPS data then we can use the Bayesian spatial model for proportion. Suppose there are n_i sample units from each small area with a binary response y_{ij} and covariates x_{ij}, $i = 1, \cdots, \ell$, $j = 1, \cdots, n_i$. The binary response y_{ij} is assumed to have an independent Bernoulli distribution with parameter π_{ij} and related linearly with the link function as

$$\log\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = \boldsymbol{x}'_{ij}\boldsymbol{\beta}_i,$$

where $\boldsymbol{\beta}_i = \boldsymbol{\delta}_i + \boldsymbol{\phi}_i$ is the regression parameter (Moura and Migon, 2002).

1.2 Approximation of Likelihood

In Bayesian statistics, if the conditional posterior or the joint posterior densities are not in simple form, parameter sampling could be computationally expensive. We have used continuous and positively skewed distributions to fit the models. The models we have fitted (see section 1.4) do not have the joint posterior densities and the conditional posterior densities in simple form. To make parameter-drawing simple and less computationally expensive, we approximate our unimodal density function by using the second-order Taylor's series, which helps with the kernel of the multivariate normal distributions for most of the parameters. Since convolution of the multivariate normal distributions is a multivariate normal distribution, this property helps if we have priors with multivariate normal distributions. For more see Nandram, Fu and Manandhar (2017). Below, we provide the important results applicable for approximation.

Lemma: Let $\pi(\boldsymbol{\tau})$ *be the unimodal density function. Then,* $\boldsymbol{\tau}$ *has an approximately multivariate normal distribution*

$$\boldsymbol{\tau} \sim N\left\{\boldsymbol{\tau}^* - H^{-1}\boldsymbol{g},\ \ -H^{-1}\right\}, \tag{1.1}$$

where $\boldsymbol{\tau}^*, \boldsymbol{g},$ *and* H *are the mode values, the gradient vector, and the Hessian matrix respectively of* $\log \pi(\boldsymbol{\tau})$.

Proof. Let $G(\boldsymbol{\tau}) = \log \pi(\boldsymbol{\tau})$ and its second-order multivariate Taylor's series approximation of $G(\boldsymbol{\tau})$ at $\tau = \tau^*$ be

$$G(\boldsymbol{\tau}) \approx G(\boldsymbol{\tau}^*) + (\boldsymbol{\tau} - \boldsymbol{\tau}^*)'\boldsymbol{g} + \frac{1}{2}(\boldsymbol{\tau} - \boldsymbol{\tau}^*)'H(\boldsymbol{\tau} - \boldsymbol{\tau}^*).$$

Then the density function $\pi(\boldsymbol{\tau})$ can be expressed as

$$\begin{aligned}
\pi(\boldsymbol{\tau}) &= e^{\log \pi(\boldsymbol{\tau})} = e^{G(\boldsymbol{\tau})} \\
&\approx e^{G(\boldsymbol{\tau}^*) + (\boldsymbol{\tau}-\boldsymbol{\tau}^*)'\boldsymbol{g} + \frac{1}{2}(\boldsymbol{\tau}-\boldsymbol{\tau}^*)'H(\boldsymbol{\tau}-\boldsymbol{\tau}^*)} \\
&= e^{G(\boldsymbol{\tau}^*) + \boldsymbol{\tau}'\boldsymbol{g} - \boldsymbol{\tau}^{*\prime}\boldsymbol{g} - \frac{1}{2}\left(-\boldsymbol{\tau}'H\boldsymbol{\tau} + 2\boldsymbol{\tau}'H\boldsymbol{\tau}^* - \boldsymbol{\tau}^{*\prime}H\boldsymbol{\tau}^*\right)} \\
&= e^{G(\boldsymbol{\tau}^*) - \boldsymbol{\tau}^{*\prime}\boldsymbol{g} + \frac{1}{2}\boldsymbol{\tau}^{*\prime}H\boldsymbol{\tau}^* - \frac{1}{2}\left(\boldsymbol{\tau}'(-H)\boldsymbol{\tau} - 2\boldsymbol{\tau}'(-H)(\boldsymbol{\tau}^* - H^{-1}\boldsymbol{g})\right)} \\
&= e^{C(\boldsymbol{\tau}^*)} e^{-\frac{1}{2}\left[\left(\boldsymbol{\tau} - (\boldsymbol{\tau}^* - H^{-1}\boldsymbol{g})\right)'(-H)\left(\boldsymbol{\tau} - (\boldsymbol{\tau}^* - H^{-1}\boldsymbol{g})\right)\right]},
\end{aligned}$$

where $C(\boldsymbol{\tau}^*)$ is a function of $\boldsymbol{\tau}^*$ only. The density function $\pi(\boldsymbol{\tau})$ can be written as

$$\pi(\boldsymbol{\tau}) \propto e^{-\frac{1}{2}\left[\left(\boldsymbol{\tau} - (\boldsymbol{\tau}^* - H^{-1}\boldsymbol{g})\right)'\,(-H)\,\left(\boldsymbol{\tau} - (\boldsymbol{\tau}^* - H^{-1}\boldsymbol{g})\right)\right]}, \tag{1.2}$$

which is the kernel for the multivariate normal distribution function with the mean vector $(\boldsymbol{\tau}^* - H^{-1}\boldsymbol{g})$ and covariance matrix $(-H^{-1})$. Therefore, $\boldsymbol{\tau}$ has approximately a multivariate normal distribution

$$\boldsymbol{\tau} \sim N\left\{\boldsymbol{\tau}^* - H^{-1}\boldsymbol{g}, -H^{-1}\right\}.$$

$\square$

Multivariate Normal Approximation Theorem

Theorem 1.2.1. *Suppose $\Delta = G(\boldsymbol{\tau})$ is the log-likelihood function of unimodal density for the given data, response y_{ij} with corresponding covariates $\boldsymbol{x}_{ij},\ i = 1, \cdots, \ell,\ j = 1, \cdots, n_i$. Let $\boldsymbol{\tau}$ can be written as $(\boldsymbol{\beta}', \boldsymbol{\nu}')'$, then the joint posterior density of the parameters can be approximated by a multivariate normal density. Furthermore, the marginal density of $\boldsymbol{\beta}$ and the conditional density of $\boldsymbol{\nu}|\boldsymbol{\beta}$ can be approximated by multivariate normal densities.*

Proof. Given the log-likelihood function of $\boldsymbol{\tau}$ and $\Delta = G(\boldsymbol{\tau})$, let us write $\boldsymbol{\tau} = \begin{pmatrix} \boldsymbol{\nu} \\ \boldsymbol{\beta} \end{pmatrix}$, with the corresponding gradient vectors $\boldsymbol{g} = \begin{pmatrix} \boldsymbol{g}_{\boldsymbol{\nu}} \\ \boldsymbol{g}_{\boldsymbol{\beta}} \end{pmatrix}$ and Hessian matrix H, evaluated at the mode $\begin{pmatrix} \boldsymbol{\nu}^* \\ \boldsymbol{\beta}^* \end{pmatrix}$, then we have

$$\boldsymbol{g} = \left(\frac{\partial \Delta}{\partial \nu_1} \ \cdots \ \frac{\partial \Delta}{\partial \nu_\ell} \ \frac{\partial \Delta}{\partial \boldsymbol{\beta_0}} \ \cdots \ \frac{\partial \Delta}{\partial \boldsymbol{\beta_p}} \right)' |_{\boldsymbol{\nu}=\boldsymbol{\nu}^*,\, \boldsymbol{\beta}=\boldsymbol{\beta}^*},$$

$$\boldsymbol{g}_{\boldsymbol{\nu}} = \left(\frac{\partial \Delta}{\partial \nu_1} \ \cdots \ \frac{\partial \Delta}{\partial \nu_\ell} \right)' |_{\boldsymbol{\nu}=\boldsymbol{\nu}^*,\, \boldsymbol{\beta}=\boldsymbol{\beta}^*}, \quad \boldsymbol{g}_{\boldsymbol{\beta}} = \left(\frac{\partial \Delta}{\partial \beta_0} \ \cdots \ \frac{\partial \Delta}{\partial \beta_p} \right)' |_{\boldsymbol{\nu}=\boldsymbol{\nu}^*,\, \boldsymbol{\beta}=\boldsymbol{\beta}^*},$$

$$H = \begin{pmatrix} \frac{\partial^2 \Delta}{\partial {\nu_1}^2} & \cdots & 0 & \frac{\partial^2 \Delta}{\partial \nu_1 \partial \boldsymbol{\beta_0}} & \cdots & \frac{\partial^2 \Delta}{\partial \nu_1 \partial \boldsymbol{\beta_p}} \\ \vdots & \ddots & \vdots & \vdots & & \\ 0 & \cdots & \frac{\partial^2 \Delta}{\partial {\nu_\ell}^2} & \frac{\partial^2 \Delta}{\partial \nu_\ell \partial \boldsymbol{\beta_0}} & \cdots & \frac{\partial^2 \Delta}{\partial \nu_\ell \partial \boldsymbol{\beta_p}} \\ \frac{\partial^2 \Delta}{\partial \nu_1 \partial \boldsymbol{\beta_0}} & \cdots & \frac{\partial^2 \Delta}{\partial \nu_\ell \partial \boldsymbol{\beta_0}} & \frac{\partial^2 \Delta}{\partial \boldsymbol{\beta_0^2}} & \cdots & \frac{\partial^2 \Delta}{\partial \boldsymbol{\beta_0} \partial \boldsymbol{\beta_p}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 \Delta}{\partial \nu_1 \partial \boldsymbol{\beta_p}} & \cdots & \frac{\partial^2 \Delta}{\partial \nu_\ell \partial \boldsymbol{\beta_p}} & \frac{\partial^2 \Delta}{\partial \boldsymbol{\beta_0} \partial \boldsymbol{\beta_p}} & \cdots & \frac{\partial^2 \Delta}{\partial \boldsymbol{\beta_p^2}} \end{pmatrix} |_{\boldsymbol{\nu}=\boldsymbol{\nu}^*, \boldsymbol{\beta}=\boldsymbol{\beta}^*}.$$

Let

$$H = -\begin{pmatrix} A_{11} & A_{12} \\ A_{12}' & A_{22} \end{pmatrix}, \tag{1.3}$$

where

$$A_{11} = -\begin{pmatrix} \frac{\partial^2\Delta}{\partial{\nu_1}^2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{\partial^2\Delta}{\partial{\nu_\ell}^2} \end{pmatrix}, A_{12} = -\begin{pmatrix} \frac{\partial^2\Delta}{\partial\nu_1\partial\boldsymbol{\beta_0}} & \cdots & \frac{\partial^2\Delta}{\partial\nu_1\partial\boldsymbol{\beta_p}} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2\Delta}{\partial\nu_\ell\partial\boldsymbol{\beta_0}} & \cdots & \frac{\partial^2\Delta}{\partial\nu_\ell\partial\boldsymbol{\beta_p}} \end{pmatrix}, A_{22} = \begin{pmatrix} \frac{\partial^2\Delta}{\partial\boldsymbol{\beta_0^2}} & \cdots & \frac{\partial^2\Delta}{\partial\boldsymbol{\beta_0}\partial\boldsymbol{\beta_p}} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2\Delta}{\partial\boldsymbol{\beta_0}\partial\boldsymbol{\beta_p}} & \cdots & \frac{\partial^2\Delta}{\partial\boldsymbol{\beta_p^2}} \end{pmatrix}.$$

From Lemma **??**, the multivariate normal approximation for the unimodal function we have

$$\begin{pmatrix} \boldsymbol{\nu} \\ \boldsymbol{\beta} \end{pmatrix} \sim N\left\{\boldsymbol{\tau}^* - H^{-1}\boldsymbol{g}, -H^{-1}\right\}, \quad \boldsymbol{\tau}^* = \begin{pmatrix} \boldsymbol{\nu}^* \\ \boldsymbol{\beta}^* \end{pmatrix},$$

where $\boldsymbol{\tau}^*$ is the approximated mode, $\boldsymbol{g}$ and H are the gradient vector and the Hessian matrix evaluated at $(\boldsymbol{\nu}^*, \boldsymbol{\beta}^*)$ respectively. Then the approximated multivariate normal distribution of $\boldsymbol{\tau}$ can be written as

$$\begin{pmatrix} \boldsymbol{\nu} \\ \boldsymbol{\beta} \end{pmatrix} \sim N\left\{\begin{pmatrix} \boldsymbol{\nu^*} \\ \boldsymbol{\beta^*} \end{pmatrix} + \begin{pmatrix} A_{11} & A_{12} \\ A_{12}' & A_{22} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{g_\nu} \\ \boldsymbol{g_\beta} \end{pmatrix}, \begin{pmatrix} A_{11} & A_{12} \\ A_{12}' & A_{22} \end{pmatrix}^{-1}\right\}, \quad \text{which can be simplified as}$$

$$\begin{pmatrix} \boldsymbol{\nu} \\ \boldsymbol{\beta} \end{pmatrix} \sim N\left\{\begin{pmatrix} \boldsymbol{\mu_\nu^*} \\ \boldsymbol{\mu_\beta^*} \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{pmatrix}\right\}, \tag{1.4}$$

where

$$\boldsymbol{\mu}_\nu^* = \boldsymbol{\nu}^* + \Sigma_{11}\boldsymbol{g_\nu} + \Sigma_{12}\boldsymbol{g_\beta}, \quad \text{and}$$

$$\boldsymbol{\mu}_\beta^* = \boldsymbol{\beta^*} + \Sigma_{12}'\boldsymbol{g_\nu} + \Sigma_{22}\boldsymbol{g_\beta}.$$

Now applying the multivariate normal theorem, we have

$$\boldsymbol{\beta}|\boldsymbol{y} \sim N\left(\boldsymbol{\mu}_\beta^*,\ \Sigma_{22}\right), \quad \text{and} \tag{1.5}$$

$$\boldsymbol{\nu}|\boldsymbol{\beta}, \boldsymbol{y} \sim N\left(\boldsymbol{\mu}_\nu^* + \Sigma_{12}\Sigma_{22}^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu}_\beta^*),\ \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}'\right). \tag{1.6}$$

□

We have used this *"multivariate normal approximation theorem"* throughout this dissertation with the same notations.

1.3 Deficiencies in Existing Models

Many models are built under the normality assumption. If the variable under study is not normal or approximately normal, a transformation of the variable is needed to meet the normality assumption. In our study we have CPS response data. Let us say we have a normality assumption for modeling CPS data, and so we need a transformation. The transformation of CPS data should remove the skewness of the data and transform the data to a bell shape. As mentioned before, the logarithmic transformation is the most popular tool used for continuous and positively skewed data. If the data follow or approximately follow the log-normal distribution, then the logarithmic-transformed data follow or approximately follow the normal distribution. However, in general we do not know the distribution except we know it is CPS data. In some cases the logarithmic transformation can make the distribution more skewed than the original data. In real studies, data could be very skewed, and standard analysis may yield invalid results (Feng et al., 2013, 2014).

If we have applied the transformation to meet the normality assumption, then the usual way to get back to the original scale is by back transformation. However, back transformation may not give a correct estimate. Here we will provide an example. Consider CPS data with the log-normal distribution. We know that if we have data from the log-normal distribution, logarithmic transformation removes skewness and gives a normal distribution. Let us say we have observed data from a log-normal distribution, $y^* \sim \mathrm{LN}(\mu, \sigma^2)$. The mean of the observed data is

$$E[y^*] = e^{\mu + \frac{\sigma^2}{2}}.$$

Let us take the log transformation of the observed data, $y = \ln(y^*)$. Now, the log-transformed observed data follow a normal distribution with mean μ_{ln}. The mean estimate of the log transformed data is

$$\hat{\mu}_{ln} = \frac{1}{n}\sum_{i=1}^{n} \log(y_i^*).$$

Transferring back to the mean we get $e^{\hat{\mu}_{ln}}$, the maximum-likelihood estimate but not an unbiased estimate of e^{μ}. However, the mean of the observed data y^* is $e^{\mu+\frac{\sigma^2}{2}}$, not e^{μ}. Thus, log-transformation cannot give correct estimates of the log-normal distribution (see Feng et al., 2013, 2014).

There is yet another problem pertinent to log-transformation. We will show an example of the non-existence of moments in the Bayesian paradigm. Let us consider a population with size N, with response variable $\boldsymbol{y}$, and we have sampled $\boldsymbol{y_s} = (y_1, \cdots, y_n)$. Let us take the log-transformation of the observed variable

$$z_i = \log(y_i), \quad i = 1, \cdots, N.$$

Consider a model and its prior

$$\begin{aligned} z_1, \cdots, z_n, z_{n+1}, \cdots, z_N|\mu, \sigma^2 &\overset{\text{iid}}{\sim} N(\mu, \sigma^2), \quad \sigma^2 > 0 \\ \pi(\mu, \sigma^2) &\propto \frac{1}{\sigma^2}. \end{aligned}$$

It gives

$$\begin{aligned} &\frac{(n-1)s^2}{\sigma^2}|\boldsymbol{z_s} \sim \text{Gamma}\left(\frac{n-1}{2}, \frac{1}{2}\right), \quad \boldsymbol{z_s} = (z_1, \cdots, z_n)', \\ &\frac{\mu - \bar{z}}{s/\sqrt{n}}|\boldsymbol{z}_s \sim t_{n-1}. \end{aligned}$$

Let us find the expected value of the response variable by integrating out parameters

$$\begin{aligned} I = E[y_i|\boldsymbol{z_s}] &= E_{\mu,\sigma^2}\left[E(y_i|\mu, \sigma^2, \boldsymbol{z_s})\right] \\ &= E_{\mu,\sigma^2}[e^{\mu+\frac{\sigma^2}{2}}|\boldsymbol{z_s}] \\ &= \int_{-\infty}^{\infty}\int_0^{\infty} e^{\mu+\frac{\sigma^2}{2}} \quad \pi(\mu|\sigma^2, \boldsymbol{z_s})\, \pi(\sigma^2|\boldsymbol{z_s})\, d\sigma^2 d\mu. \end{aligned}$$

Since $\frac{\sigma^2}{2} > 0$,

$$\begin{aligned} I &\geq \int_{-\infty}^{\infty}\int_0^{\infty} e^{\mu} \quad \pi(\mu|\sigma^2, \boldsymbol{z_s})\, \pi(\sigma^2|\boldsymbol{z_s})\, d\sigma^2 d\mu \\ &= \int_{-\infty}^{\infty} e^{\mu} \int_0^{\infty} \quad \pi(\mu|\sigma^2, \boldsymbol{z_s})\, \pi(\sigma^2|\boldsymbol{z_s})\, d\sigma^2 d\mu \\ &= \int_{-\infty}^{\infty} e^{\mu} \quad t_{n-1}\left(\frac{\mu - \bar{z}}{s/\sqrt{n}}|\boldsymbol{z_s}\right) d\mu = \infty. \end{aligned}$$

Thus, the expected value of y_i does not exist because the moment generating function of the Student's t distribution does not exist. This is indeed problematic for inference.

1.4 Distributions Used in Model Building

We would like to have a model for a CPS data without log transformation. To facilitate the modeling, we need a continuous and positively skewed probability-density function. Which continuous and positively skewed density should be chosen? In general, we do not know the distribution of the responses. If we have very little knowledge about the distribution of the response variable, it would be better to choose a generalized distribution for the model. However, a generalized distribution has more parameters and there are difficulties in sampling the parameters.

In this dissertation, we start our model with a simple one-parameter density function and move to a generalized four-parameter density function. First, we start the model with the exponential distribution. Second, the gamma distribution, which is more generalized than the exponential distribution. Third, the generalized gamma distribution, which is more generalized than the gamma distribution. Finally the GB2 distribution, which is a scale mixture of the two generalized gamma distributions. In total we have four density functions in hierarchical order, from the simple exponential to the generalized GB2 density function. Since the GB2 is the most generalized density function, we consider three special cases: first, a mixture of the exponential and the gamma distributions; second, a mixture of two gamma distributions; and third, a mixture of two generalized gamma distributions. We choose the distribution for modeling our CPS data according to whether the responses are noisy or not.

In aggregate we have six density functions that have been discussed in this dissertation: the exponential, the gamma, the generalized gamma, and the three special cases of the GB2 density. For each density we will fit two models, one without random area effects and the other with random area effects. Overall, this dissertation explains twelve models, six models without random area effects and six models with random area effects.

1.4.1 Modeling with standard distributions

The standard distribution in this dissertation refers to the generalized gamma distribution or its special cases, which are not a mixture of two generalized gamma distributions as the GB2 distribution is. For standard distribution models we have chosen three distributions: the exponential, the gamma, and the generalized gamma distributions. We fit the standard distribution for modeling noiseless CPS data.

Exponential Model: We assume the response variable $y|\lambda$ has the exponential distribution with rate λ and fit models, with and without random area effects

$$\begin{aligned} Y|\lambda &\sim \text{Exp}\left(\lambda\right), \quad \lambda > 0,\ y > 0 \\ f(y|\lambda) &= \lambda e^{-\lambda y}. \end{aligned} \tag{1.7}$$

Gamma Model: We assume the response variable $y|\alpha, \lambda$ has the gamma distribution with shape α and rate λ and fit models, with and without random area effects

$$\begin{aligned} y|\alpha,\lambda &\sim \text{Gamma}\left(\alpha,\lambda\right), \quad \lambda,\alpha > 0,\ y > 0 \\ f(y|\alpha,\lambda) &= \frac{e^{-\lambda y}\ y^{\alpha-1}}{\Gamma(\alpha)}\ \lambda^{\alpha}. \end{aligned} \tag{1.8}$$

Generalized Gamma Model: We assume the response variable $y|\alpha, \lambda, \gamma$ has the generalized gamma distribution with shape α, γ and rate λ and fit models, with and without random area effects

$$\begin{aligned} y|\alpha,\lambda,\gamma &\sim \text{GGamma}\left(\alpha,\lambda,\gamma\right), \quad \lambda,\alpha,\gamma > 0,\ y > 0, \\ f(y|\alpha,\lambda,\gamma) &= \gamma\ \frac{e^{-\lambda y^{\gamma}}\ y^{\alpha\gamma-1}}{\Gamma(\alpha)}\ \lambda^{\alpha}. \end{aligned} \tag{1.9}$$

In these models, we introduce covariates through their rate parameters. If there are no random area effects in the model, then $\lambda = e^{-\boldsymbol{x}'\boldsymbol{\beta}}$. If there are random area effects, then $\lambda = e^{-(\boldsymbol{x}'\boldsymbol{\beta}+\nu)}$, where ν is the random area effect parameter. Parameters are drawn using the Metropolis Hastings algorithm and the grid sampling method. Predictions of the response variable have been drawn from their respective distributions.

1.4.2 Modeling with GB2 Distribution

The GB2 is a special case of the generalized beta-F distribution. Let a random variable Y have a beta distribution, $Y \sim \text{Beta}(\alpha, \phi)$. Its cumulative distribution function (CDF) is given by

$$G_Y(y) = \frac{1}{\mathrm{B}(\alpha,\phi)} \int_0^y t^{\alpha-1}(1-t)^{\phi-1}dt, \qquad \alpha,\phi > 0,\ 0 < y < 1,$$

where $\mathrm{B}(\alpha,\phi) = \frac{\Gamma(\alpha)\Gamma(\phi)}{\Gamma(\alpha+\phi)}$ is the beta function. We note that G(.) has both its domain and range (0, 1). Replacing the upper limit of the integration by any cumulative distribution function F(y) and differentiating, we get the generalized beta-F probability density function (Sepanski and Kong, 2007).

$$G_F(y) = \frac{1}{\mathrm{B}(\alpha,\phi)} \int_0^{F(y)} t^{\alpha-1}(1-t)^{\phi-1}dt, \qquad \alpha,\phi > 0,\ 0 < y < \infty, \tag{1.10}$$

$$g_F(y) = \frac{1}{\mathrm{B}(\alpha,\phi)} f(y)\,[F(y)]^{\alpha-1}\,[1-F(y)]^{\phi-1}, \qquad \alpha,\phi > 0. \tag{1.11}$$

Since CDF, F(.) can come from any density function, this generalized beta-F is very rich. This family of distribution was first introduced by Singh et al. (1988). Note, F is not to be confused with the F-distribution. Actually this holds for any random variable $y \in (-\infty, \infty)$. Because we are interested in size data we restrict $y \in (0, \infty)$. We will consider two special cases of generalized beta-F distribution. If $F(y) = (\frac{y}{\theta})^{\gamma}$, we get the generalized beta distribution of the first kind (GB1)

$$g_F(y) = \frac{|\gamma|\ y^{\gamma\alpha-1}\left[\ 1-(\frac{y}{\theta})^{\gamma}\ \right]^{\phi-1}}{\theta^{\gamma\alpha}\ \mathrm{B}(\alpha,\phi)}, \qquad 0 < y \le \theta. \tag{1.12}$$

If we consider $F(y) = 1 - [\ 1 + (\frac{y}{\theta})^{\gamma}\ ]^{-1}$, we get the generalized beta distribution of the second kind (GB2)

$$g_F(y) = \frac{|\gamma|\ y^{\gamma\alpha-1}}{\theta^{\gamma\alpha}\ \mathrm{B}(\alpha,\phi)\left[\ 1+(\frac{y}{\theta})^{\gamma}\ \right]^{\alpha+\phi}}; \qquad y > 0. \tag{1.13}$$

By reparameterizing the parameters, this GB2 density can be written as:

$$g_F(y) = \ = \frac{|\gamma|}{\theta^{\alpha}\mathrm{B}\left(\frac{\alpha}{\gamma},\frac{\phi}{\gamma}\right)} \frac{y^{\alpha-1}}{\left(1+\left(\frac{y}{\theta}\right)^{\gamma}\right)^{\frac{\alpha+\phi}{\gamma}}}, \qquad y,\alpha,\phi > 0. \tag{1.14}$$

We use GB2 to model size (positive values) data. GB2 density can also be expressed as a scale mixture of two generalized gamma distributions. We exploit this property in model building. Let the probability density function of the response variable $y|\alpha, \lambda, \gamma$ and the probability density function of its rate parameter $\lambda|\phi, \theta, \gamma$ both have the generalized gamma distribution

$$y|\alpha, \lambda, \gamma \sim \text{GGamma}(\alpha, \lambda, \gamma), \text{ and}$$
$$\lambda|\phi, \theta, \gamma \sim \text{GGamma}(\phi, \theta, \gamma).$$

We note that we have same γ parameter in both distributions. Mixing the generalized gamma density function of y with the generalized gamma density function of its rate parameter λ and integrating out λ gives the GB2 density with four parameters $\alpha, \phi, \gamma, \text{and}, \theta$

$$\begin{aligned} f(y|\alpha, \phi, \theta) &= \int_0^\infty f(y|\lambda, \alpha, \gamma) \;\; g(\lambda|\theta, \phi, \gamma) \, d\lambda \\ &= \frac{\gamma}{\mathrm{B}(\frac{\alpha}{\gamma}, \frac{\phi}{\gamma})} \frac{y^{\alpha-1}}{\theta^\alpha (1 + \left(\frac{y}{\theta}\right)^\gamma)^{\frac{\alpha+\phi}{\gamma}}}, \quad \theta, \alpha, \phi > 0, \, y > 0. \end{aligned} \tag{1.15}$$

We have derived GB2 as a mixture of two generalized gamma densities. We note here that the rate parameter λ of the response variable $\boldsymbol{y}$ has been integrated out and the new shape and rate parameters, (ϕ, θ), introduced from the distribution of λ. Therefore, in the GB2 density function λ is hidden and it has one more fold of distribution than the generalized gamma distribution. The distribution of this rate parameter contributes in adding noises in the true response values. As an illustrative example let us say the true value of response is μ_i but we observed response y_i as

$$y_i = \mu_i + e_i, \quad e_i \sim N(0, \sigma^2),$$

therefore the observed response y_i is noisy. Similarly, the distribution of the λ adds noises in the true values. We use this property of GB2 to model the noisy responses.

The generalized gamma density is itself a generalized density with three parameters: two shape parameters and one rate parameter. We choose three special cases of GB2 distribution for modeling the CPS data. Let a random variable Y have the generalized gamma distribution, $Y|\alpha, \lambda, \gamma \sim \text{GGamma}(\alpha, \lambda, \gamma)$. In the generalized gamma distribution, if $\gamma = 1$ then we get the gamma distribution $Y|\alpha, \lambda \sim \text{Gamma}(\alpha, \lambda)$. If $\gamma = \alpha = 1$, then we get the exponential distribution $Y|\lambda \sim \text{Exp}(\lambda)$. Below we show GB2 models as a mixture of the exponential and the gamma distributions, a mixture of two gamma distributions, and a mixture of the two generalized gamma distributions.

Exponential-Gamma Mixture Model: The simplest GB2 density we are using for a model is a mixture of the exponential and the gamma distributions. Let us consider if the response variable $Y|\lambda$ has the exponential distribution, and its rate parameter $\lambda|\alpha, \theta$ has the gamma distribution. Mixing these two distributions we get GB2 density

$$f(y|\alpha, \theta) = \frac{\alpha}{\theta(1 + \frac{y}{\theta})^{\alpha+1}}\ , \quad \alpha, \theta > 0. \tag{1.16}$$

We note that moments do not exist for a mixture of the two exponential distributions. Let the response variable have an exponential distribution $Y|\lambda \sim \text{Exp}(\lambda)$, and its rate parameter λ have the exponential distributions $\lambda|\theta \sim \text{Exp}(\theta)$

$$f(y|\lambda) = \lambda e^{-\lambda y},\ \lambda > 0, \quad f(\lambda|\theta) = \theta e^{-\theta\lambda},\ \theta > 0.$$

Mixing these two exponential densities and integrating out λ we get GB2 density

$$\begin{aligned} f(y|\theta) = \theta \int_{\lambda} \lambda e^{-\lambda y}\ e^{-\theta\lambda} d\lambda &= \frac{1}{\theta(1 + \frac{y}{\theta})^2}\ , \quad \theta > 0, \\ E\left[Y^k\right] &= E_{\lambda}\left[E\left[Y^k|\lambda\right]\right] = \Gamma(k+1)\ \theta \int_0^{\infty} e^{-\theta\lambda}\lambda^{-k} d\lambda \\ &= \Gamma(k+1)\ \theta \left\{\int_0^1 e^{-\theta\lambda}\lambda^{-k} d\lambda + \int_1^{\infty} e^{-\theta\lambda}\lambda^{-k} d\lambda\right\}. \end{aligned}$$

The integral $\int_0^{\infty} e^{-\theta\lambda}\lambda^{-k} d\lambda$ diverges for all k. Therefore moments do not exist for the mixture of the two exponential distributions.

Mixture of two Gamma GB2 Model: Let the response variable have the gamma distribution $Y|\alpha, \lambda \sim \text{Gamma}(\alpha, \lambda)$ and its rate parameter have the gamma distribution $\lambda|\phi, \theta \sim \text{Gamma}(\phi, \theta)$. Mixing these two gamma densities and integrating out λ, we get the GB2

$$\begin{aligned} f(y|\alpha,\phi,\theta) &= \frac{y^{\alpha-1}\theta^{\phi}}{\Gamma(\alpha)\Gamma(\phi)}\int_{\lambda} e^{-(\theta+y)\lambda} \ \ \lambda^{\alpha+\phi-1}d\lambda \\ &= \frac{y^{\alpha-1}}{\mathrm{B}(\alpha,\phi)}\,\frac{1}{\theta^{\alpha}(1+\frac{y}{\theta})^{\alpha+\phi}}, \quad \theta,\alpha,\phi>0. \end{aligned} \tag{1.17}$$

Its k^{th} moment is given by

$$E[Y^k|\alpha,\phi,\theta] = \frac{\Gamma(\alpha+k)}{\Gamma(\alpha)}\frac{\Gamma(\phi-k)}{\Gamma(\phi)}\theta^k. \tag{1.18}$$

For variance to exist in this density, we need $\phi > 2$. If we consider α and ϕ as two distinct shape parameters, they are not identifiable (see below). We consider that two rate parameters, α and ϕ, are related linearly as, $\phi = \alpha + 2$. This also allows the variance to exist. Considering this linear relationship between two shape parameters, we have the GB2 density function from the mixture of two gamma distributions as

$$f(y|\alpha,\theta) = \frac{y^{\alpha-1}}{\mathrm{B}(\alpha,\alpha+2)}\,\frac{1}{\theta^{\alpha}(1+\frac{y}{\theta})^{2(\alpha+1)}}, \quad \theta,\alpha>0. \tag{1.19}$$

1.4.3 Non-identifiable Parameters in GB2 Distribution

Let us say we have n independent samples from the gamma distribution, $Y_i|\lambda_i,\alpha \sim \text{Gamma}(\lambda_i,\alpha)$. We would like to find a maximum likelihood estimate (MLE) for the parameters $\alpha, \lambda_i, i = 1,\cdots,n$, the likelihood function given as

$$f(\boldsymbol{y}|\alpha,\boldsymbol{\lambda}) = \prod_{i=1}^{n}\frac{e^{-\lambda_i y_i}\ y^{\alpha-1}}{\Gamma(\alpha)}\ \lambda_i^{\alpha}, \quad \lambda_i,\alpha>0,\ i=1,\cdots,n.$$

The log-likelihood function is

$$\Delta = \sum_{i=1}^{n}[-\lambda_i y_i + (\alpha-1)log(y_i) + \alpha log(\lambda_i)] - nlog\Gamma(\alpha).$$

The maximum likelihood estimator (MLE) for λ_i, is $\hat{\lambda}_i = \frac{\alpha}{y_i},\ i = 1,\cdots,\ell$. Substituting MLE of λ_i in the log-likelihood function gives us

$$\Delta = \sum_{i=1}^{n}[-\alpha + (\alpha-1)log(y_i) + \alpha log(\alpha/y_i)] - nlog\Gamma(\alpha).$$

Taking the partial derivative with respect to α gives

$$\frac{\partial \Delta}{\partial \alpha} = nlog(\alpha) - n\Psi(\alpha),$$

where $\Psi(\alpha) = \frac{d}{d\alpha} log\Gamma(\alpha)$. Setting $\frac{\partial \Delta}{\partial \alpha} = 0$, we have $log(\alpha) = \Psi(\alpha)$. It has no solution. Therefore, if each response y_i has its parameter $\lambda_i, i = 1, \cdots, n$, then the parameters (α, λ_i) together are not identifiable.

Non-Identifiable Illustration Here is another illustration on non-identifiable parameters. Let us consider the hierarchical Bayesian model

$$y_i|\mu_i \overset{\text{ind}}{\sim} N(\mu_i, \sigma^2), \quad \mu_i|\theta, \delta^2 \overset{\text{iid}}{\sim} N(\theta, \delta^2), \quad i = 1, \cdots, n.$$

Integrating out μ_i from the posterior density function, we get the marginal distribution function

$$y_i|\theta, \sigma^2, \delta^2 \sim N(\theta,\, \sigma^2 + \delta^2).$$

Here, θ and $\sigma^2 + \delta^2$ are identifiable. However, distinct σ^2 and δ^2 are not identifiable. This type of models is extensively seen in *spatial-temporal data* (Cressie and Wikle, 2015) in which σ^2 is assumed to be known (instrumental error).

Mixture of two Generalized Gamma GB2 Model: Let the response variable have the generalized gamma distribution, $Y|\alpha, \lambda, \gamma \sim \text{GGamma}(\alpha, \lambda, \gamma)$ and let its rate parameter also have the generalized gamma distribution, $\lambda|\phi, \theta, \gamma \sim \text{GGamma}(\phi, \theta, \gamma)$. Mixing these two distributions and integrating out λ, we have the following GB2 density

$$\begin{aligned} f(y|\alpha, \phi, \theta, \gamma) &= \gamma^2 \frac{y^{\alpha-1}\theta^{\phi}}{\Gamma(\frac{\alpha}{\gamma})\Gamma(\frac{\phi}{\gamma})} \int_{\lambda} e^{-(\theta^{\gamma}+y^{\gamma})\lambda^{\gamma}}\, \lambda^{\alpha+\phi-1}\, d\lambda, \\ &= \frac{\gamma\, y^{\alpha-1}}{\mathrm{B}\left(\frac{\alpha}{\gamma}, \frac{\phi}{\gamma}\right)} \frac{1}{\theta^{\alpha}(1+(\frac{y}{\theta})^{\gamma})^{\frac{\alpha+\phi}{\gamma}}}, \quad \theta, \alpha, \phi, \gamma > 0. \end{aligned} \tag{1.20}$$

Its k^{th} moment is

$$E[Y^k|\alpha, \phi, \theta, \gamma] = \frac{\Gamma\left(\frac{\alpha+k}{\gamma}\right)}{\Gamma\left(\frac{\alpha}{\gamma}\right)} \frac{\Gamma\left(\frac{\phi-k}{\gamma}\right)}{\Gamma\left(\frac{\phi}{\gamma}\right)} \theta^k. \tag{1.21}$$

As before, in the mixture of the two gamma distributions, we need $\phi > 2$ for the variance to exist. We assume that the two shape parameters α and ϕ are related linearly as, $\phi = \alpha + 2$. Then the GB2 density can be written as

$$f(y|\alpha, \theta, \gamma) = \frac{\gamma y^{\alpha-1}}{\mathrm{B}(\frac{\alpha}{\gamma}, \frac{\alpha+2}{\gamma})} \frac{1}{\theta^{\alpha}(1 + (\frac{y}{\theta})^{\gamma})^{\frac{2(\alpha+1)}{\gamma}}}, \quad \theta, \alpha, \gamma > 0. \tag{1.22}$$

1.5 Family of Poverty Measures

Let E_{ij} be the welfare measure for the j^{th} unit of the i^{th} area, $i = 1, \cdots, A,\ j = 1, \cdots, N_i$. The family of poverty measures for small area i given the predetermined poverty threshold $z > 0$ (Foster, Greer, & Thorbecke, 1984) is given by

$$P_{\alpha i} = \frac{1}{N_i} \sum_{j=1}^{N_i} \left(\frac{z - E_{ij}}{z} \right)^{\alpha} I\left(E_{ij} < z\right), \qquad \alpha \geq 0, \quad i = 1, \cdots, A,$$

where $I(E_{ij} < z)$ is an indicator function, equals 1 if $(E_{ij} < z)$ else zero. Here $(z - E_{ij})$ is the consumption shortfall of the j^{th} unit in the i^{th} area.

If $\alpha = 0, \mathrm{P}_{0i}$ gives poverty incidence, the proportion of poor,

$$P_{0i} = \frac{N_{0i}}{N_i}, \quad \text{where} \quad N_{0i} = \sum_{j=1}^{N_i} I(E_{ij} < z).$$

If $\alpha = 1$, P_{1i} gives the poverty gap, the intensity of poverty. It is thought as the cost required to eliminate poverty relative to the poverty line, since it shows the amount needed to overcome poverty. If $\alpha = 2$, P_{2i} gives the poverty severity. The larger the value of the parameter α the greater emphasis given to the poorest poor. It should be clear that, it is easier to estimate P_{0i} and much more difficult to estimate P_{1i} and P_{2i}.

1.6 Application

We have applied our models to the CPS consumption data from the second Nepal Living Standards Survey (NLSS-II), conducted in the years 2003–2004. NLSS-II follows the World Bank's Living Standards Measurement Survey methodology (LSMS), which was successfully applied previously in many parts of the world. It is an integrated survey which covers samples from a whole country and runs throughout the year. The main objective of the NLSS-II is to collect data from Nepalese households and provide information to monitor progress about national living standards. The NLSS-II collects information in many fields like demographics, housing, education, access to facilities, food expenditures, non-food expenditures, and health.

1.6.1 Sample Design

The sampling design of the NLSS-II is two-stage stratified sampling. The nation was stratified into six strata: (1) Mountains, (2) Kathmandu Valley Urban, (3) Other Hills Urban, (4) Hills Rural, (5) Terai Urban, and (6) Terai Rural. In the first stage, 326 PSUs were selected from six strata using the probability proportional to size (PPS) sampling with the number of households used as a measure of size. In the second stage, 3,912 households were selected by the systematic sampling (NLSS-II report Vol 1, 2004). These 3,912 households' data are available in the NLSS-II data set. There are some other PSUs in NLSS-II where enumeration was not successful because of the Maosit insurgency at that time and data is not available.

Table 1.1: Distribution of PSUs and households in the sample and the sample frame [NLSS-II].

Stratum		**NLSS-II sample**		**Sample frame**		
Stratum	**Stratum Name**	**PSUs**	**Hhlds**	**PSUs**	**Hhlds**	**Population**
1	Mountains	32	384	4,540	321,680	1,690,263
2	Kathmandu Valley Urban	34	408	537	227,637	1,021,007
3	Other Hills Urban	28	336	382	161,922	728,039
4	Hills Rural	96	1,152	17,824	1,619,440	8,522,460
5	Terai Urban	34	408	545	294,751	1,508,102
6	Terai Rural	102	1,224	12,239	1,686,317	9,744,810
	Total	**326**	**3,912**	**36,067**	**4,311,747**	**23,214,681**

Table 1.1 shows the distribution of PSUs and households by stratum in NLSS-II. We can see that only about 0.091% of households and only 0.904% of PSUs were sampled. NLSS-II was designed to give estimates only at stratum level or larger areas than stratum. It cannot give estimates in small areas (example districts or VDC/ municipalities) because the sample sizes are too small.

NLSS-II is a two stage stratified random sampling. The country is divided into six strata, $s = 1, \cdots, 6$. In the first stage the wards were selected with PPS sampling. For the larger wards, they created subwards, especially in urban wards; and for small wards they have merged wards. At an intermediate stage, the subwards are selected by PPS sampling within the selected ward. At the second stage, 12 households are selected by systematic random sampling. In NLSS-II, the PSUs can be considered as the wards, however sometimes the PSU could be the subward or the unions of the wards. The overall probability of selection for an household in the strutum is

$$k \frac{N_i}{\sum_i N_i} \frac{K_{ij}}{\sum_j K_{ij}} \frac{12}{K_{ij}^*}$$

where k is the number of PSUs selected in the stratum

N_i is the number of household in the i^{th} ward

$\sum_{i \in \text{stratum}} N_i$ is the total number of households in the stratum

K_{ij} is the number of dwellings quick-counted in the j^{th} subward and i^{th} ward

$\sum_{j \in \text{subward}} K_{ij}$ is the total number of dwellings quick-counted in the j^{th} subward and i^{th} ward

K_{ij}^* is the number of households counted in the listing stage.

The grossing up factor is the inverse of this probability, which can be written as

$$\frac{\sum_i N_i}{12\,k} \frac{\sum_j K_{ij}}{N_i} \frac{K_{ij}^*}{K_{ij}}$$

If we have $\frac{\sum_j K_{ij}}{N_i}$ and $\frac{K_{ij}^*}{K_{ij}}$ both equal to one then we have equal selection probability $\frac{\sum_i N_i}{12\,k}$ for all households within the stratum. We assume that at the PSU level the household weights are equal.

1.6.2 Quality of Data

To maintain the quality of data, a complete household listing operation was undertaken in each selected PSU during March–May of 2002 about a year prior to the survey. Systematic sample selection of households was done in the central office. The field staff consisted of supervisors, enumerators, and data-entry operators. Female interviewers were hired to interview female respondents for sensitive questions, which are related to women, such as their marriage and maternity history. Data collection was carried out April 2003–April 2004 in an attempt to cover a complete cycle of agricultural activities and health-related questions and to capture seasonal variations in different variables. The process was completed in three phases. Data entry was done in the field. A custom data-entry program with consistency checking was developed for this survey. There was consistency checking for each questionnaire linked between sections. All errors and inconsistencies were resolved in the field. Data were collected throughout the year (NLSS-II report Vol 1, 2004).

1.6.3 Response Variable

We have a response variable, CPS per capita consumption from NLSS-II. In the living standards survey the welfare response variable per capita consumption is the aggregate of all food and non-food items consumed in the past twelve months. The per capita consumption data are available for all 3,912 households enumerated in the NLSS-II, 2003–2004 survey. In NLSS-II, expenditures on food items have been collected separately for (a) home production, (b) food purchases, and (c) food in-kind.

For each home production and food purchases item, it collects the number of months consumed in the past 12 months and quantity (with unit) consumed in a typical month during which the food item is consumed. For home production items, it records the amount that the household has to spend in the market to buy the food quantity consumed in a typical month. For food purchases, it records the amount the household normally spends to purchase this quantity. For food in-kind items, it records the total value of the food consumed over the past 12 months. For non-food items, it records goods purchased or in-kind received in money value for the past 30 days and past 12 months (NLSS-II report Vol 2, 2004).

In a living standards survey, the responders had to recall all kinds of consumptions in monetary value throughout the whole reference year. In addition, for each food item the respondent had to recall the number of months consumed and quantity consumed in the typical month, then evaluate its market value at that time. It could also be possible that there could be bias of reporting excessive quantity and excessive money values of food and non-food consumed by some household or reporting less quantity and less money values by other households. Hence, there could be the possibility of introducing noise in these kinds of data.

We build models assuming response variable is noiseless or noisy as discussed above. We build the models for noiseless responses in Chapter 2, where we fit the standard distributions: the exponential, the gamma, and the generalized gamma distributions. In Chapter 3 we assume that responses are recorded with noise and we fit the GB2 distribution.

We select the best fitted models assuming the responses are noiseless or noisy. From the best selected model from each chapter we predict per capita consumption based on the 2001 census data. To calculate the poverty indicators we compare predicted per capita consumption against the national poverty line for Nepal of 7,696 rupees per year in average 2003 Nepalese rupees, adjusted for spatial price variation. This is the same threshold Haslett et al. (2006) used for calculating poverty indicators. Finally we provide SAE of poverty indicators in Chapter 4.

1.6.4 Covariates

We chose nine relevant covariates which can influence welfare status, and per capita consumption from the NLSS-II survey for modeling. Obviously these covariates are also available in the 2001 census data. They are (i) "Household size" (*hhsize*), (ii) "proportion of kids aged 0 - 6 in the household" (*skids6*), (iii) "proportion of kids aged 7 - 14 in the household" (*skids714*), (iv) "abroad migrant" (*remtab*), (v) "House temporary" (*hutype3*), (vi) "House owned" (*huown2*), (vii) "proportion of households with cooking fuel LP/gas in Ward" (*ckfuel3w*), (viii) "proportion of household with land-owning females in municipality/VDC" (*pflandv*), and (ix) "proportion of kids 6-16 attending school in municipality/VDC" (*pschv*) from NLSS-II 2003–2004. These covariates have a moderate correlation with the response variable.

Six covariates of these are directly related with the household: *hhsize, skids6, skids714, remtab, hutype3,* and *huown2*. Size of household and proportion of children in different age group have influence on expenditure and consumption. Any household member as abroad migrant indicates the sources of remittance. Covariate *"hutype3"* indicates a temporary type of house; there are three types of houses according to the construction material of the outside walls of the house: permanent, semi-permanent, and temporary. The house owned variable is a binary variable that indicates whether the household has its own house to live in or not. Covariate *"huown2"* indicates that the household has their own house to live in.

The other three area level variables are *ckfuel3w, pflandv,* and *pschv.* Around 2003–2004 in Nepal very few rural areas used LP/gas as cooking fuel while urban households did. LP/gas is expensive compared to other sources of fuel. Covariate *"pflandv"* indicates the proportion of households with a female owning land in the Municipality/VDC. The proportion of children aged 6–16, *pschv*, who are supposed to be in school indicates the awareness of the community and strength of the future.

1.6.5 Prediction

We develop our models in the sample survey and select the best fitted models assuming noiseless and noisy responses. We apply our selected models to predict the responses in the large survey, census data in our application. To link the sample survey and large survey we carry the parameters from the fitted model to the large survey. We predict the responses in the large survey data. Prediction are done from their respective assumed distribution of the responses in the model. After prediction of the responses to each unit in the large survey, we provide the small area estimations.

As an illustrative example, consider the population of size N with

$$y_i|\boldsymbol{\beta}, \sigma^2, \boldsymbol{x}_i \overset{\text{ind}}{\sim} N(\boldsymbol{x}_i'\boldsymbol{\beta}, \sigma^2), \quad i = 1, \cdots, N,$$

where $\boldsymbol{y}$ is the response variable and $\boldsymbol{x}$ the covariates. We are interested in $\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} y_i$. We draw a sample of size n from the population and build a model

$$y_i|\boldsymbol{\beta}, \sigma^2, \boldsymbol{x}_i \overset{\text{ind}}{\sim} N(\boldsymbol{x}_i'\boldsymbol{\beta}, \sigma^2), \quad i = 1, \cdots, n$$
$$\pi(\boldsymbol{\beta}, \sigma^2) \qquad \text{with some priors.}$$

We draw the sets of M samples $h = 1, \cdots, M$, from the posterior distribution function $\pi(\boldsymbol{\beta}, \sigma^2|\boldsymbol{y})$. For each set of sample $(\boldsymbol{\beta}, \sigma^2)$, we predict the responses, $y_i^{(h)},\ h = 1, \cdots, M,\ i = 1, \cdots, N$ in the population data. The estimated mean response is $\bar{Y}^{(h)} = \frac{1}{N}\sum_{i=1}^{N} {y_i}^{(h)}$, and the inference can be made for $\bar{Y}$.

1.7 Plan for Dissertation

This dissertation has three additional chapters. In Chapter 2, we build models assuming the responses are noiseless and fit the standard distributions: the exponential, the gamma and the generalized gamma model for CPS data. First, we present models without random area effects and then we present models with random area effects. In Chapter 3, we build models assuming the responses are noisy and fit the GB2 distribution. We choose three special cases of the GB2 distribution for modeling CPS data. In the first part, we explain models without random area effects, and then we explain the model with random area effects. In Chapter 4, we apply our best fitted models, assuming noiseless and noisy responses, to the population census data and we make concluding remarks. In this chapter we also show the simulation studies of our best selected models to assure the quality of the selected models.

Chapter 2

Models for Noiseless Responses

In this chapter, we assume that observed responses are free of noises, and we fit standard distributions without logarithmic-transformation to the responses. In this dissertation, the name "standard distributions" is given for those distributions which are not the mixture of two generalized gamma (GB2) distributions. We present three standard distributions in hierarchical order from simple to generalized distributions: exponential, gamma, and generalized-gamma distribution, support $(0, \infty)$. We fit two models for each distribution, one without random area effects and another with random area effects. Thus, in this chapter we discuss six hierarchical Bayesian models.

The joint posterior density and the conditional posterior density functions of these models are not in simple form, so we use the second-order Taylor's series approximation to facilitate the sampling procedures. The Taylor's series approximation helps us to approximate a unimodal density function by providing a multivariate-normal approximated distribution.

We apply our models to the CPS welfare consumption data with nine covariates from NLSS-II: (i) "Household size" (*hhsize*), (ii) "proportion of kids aged 0 - 6 in the household" (*skids6*), (iii) "proportion of kids aged 7 - 14 in the household" (*skids714*), (iv) "abroad migrant" (*remtab*), (v) "House temporary" (*hutype3*), (vi) "House owned" (*huown2*), (vii) "proportion of households with cooking fuel LP/gas in Ward" (*ckfuel3w*), (viii) "proportion of household with land-owning females in municipality/VDC" (*pflandv*), and (ix) "proportion of kids 6-16 attending school in municipality/VDC" (*pschv*). These nine covariates are generated in both the NLSS-II and the 2001 population census for the purpose of the

SAE and their consistencies have been checked prior to modeling.

We developed models using NLSS-II household survey data and applied our models to 2001 Census data for SAE. We sampled parameters using the grid sampling method and the Metropolis–Hastings (MH) algorithm. For the MH algorithm we have a multivariate t-distribution with d degrees of freedom as a proposal density function. For all models fitted in this chapter, we take 3 degrees of freedom for proposal distribution, (d=3). We have predicted the response variable in the census data and presented the results of SAE. We calculated the conditional predictive distribution (CPO) and the summary statistics logarithm of the pseudo-marginal likelihood (LPML) for model comparisons.

Notation

Consider sample data with n observations, response variable $\boldsymbol{y}_{n\times 1}$ and covariate $\boldsymbol{x}_{n\times p}$. For a model without random area effects, we introduce covariates through the rate parameter, writing the rate parameter as $e^{-\boldsymbol{x}_i'\boldsymbol{\beta}}$, $i = 1, \dots, n$.

To build models with random area effects, we have ℓ small areas, $i = 1, \cdots, \ell$, and each small area has $j = 1, \cdots, n_i$ observations. Let y_{ij} and $\boldsymbol{x}_{ij}$, $i = 1 \cdots, \ell$, $j = 1 \cdots, n_i$ denote the response variable and the corresponding covariates in the i^{th} area and j^{th} observation. We introduce covariates through the rate parameter, writing the rate parameter as $e^{-(\boldsymbol{x}_{ij}'\boldsymbol{\beta}+\nu_i)}$.

2.1 Exponential Model without Random Area Effects

We assume $y_i|\beta$, $i = 1, \cdots, n$, is a random sample from an exponential distribution with rate $e^{-\boldsymbol{x}_i'\boldsymbol{\beta}}$. The hierarchical Bayesian exponential model with a non-informative prior for $\boldsymbol{\beta}$ is

$$\begin{aligned} & y_i|\boldsymbol{\beta} \overset{\text{ind}}{\sim} \text{Expo}(e^{-\boldsymbol{x}_i'\boldsymbol{\beta}}), \quad \lambda_i = e^{-\boldsymbol{x}_i'\boldsymbol{\beta}}, i = 1, \cdots, n, \\ & \pi(\boldsymbol{\beta}) \propto 1, \end{aligned} \tag{2.1}$$

with the likelihood function

$$
\begin{aligned}
f(\boldsymbol{y}|\boldsymbol{\beta}) &= \prod_{i=1}^{n} e^{-\boldsymbol{x_i'\beta}}\, e^{-y_i\, e^{-\boldsymbol{x_i'\beta}}} \\
&= e^{-\sum_{i=1}^{n}\left(\boldsymbol{x_i'\beta} + y_i\, e^{-\boldsymbol{x_i'\beta}}\right)}. \qquad (2.2)
\end{aligned}
$$

Combining the likelihood in (2.2) and the priors in (2.1) via Bayes' theorem, given the sample data, we get the joint posterior density of $\boldsymbol{\beta}$ as

$$
\begin{aligned}
\pi(\boldsymbol{\beta}|\boldsymbol{y}) &\propto f(\boldsymbol{y}|\boldsymbol{\beta})\, \pi(\boldsymbol{\beta}) \\
&= e^{-\sum_{i=1}^{n}\left(\boldsymbol{x_i'\beta} + y_i\, e^{-\boldsymbol{x_i'\beta}}\right)}. \qquad (2.3)
\end{aligned}
$$

Let the log-likelihood function be

$$
G(\boldsymbol{\beta}|\boldsymbol{y}) = -\sum_{i=1}^{n}\left(\boldsymbol{x_i'\beta} \,+\, y_i\, e^{-\boldsymbol{x_i'\beta}}\right).
$$

For notational simplicity, let us write G for the log-likelihood function. Then its first- and second-order partial derivatives with respect to $\boldsymbol{\beta}$ are

$$
\frac{\partial G}{\partial \boldsymbol{\beta}} = -\sum_{i=1}^{n}\left(\boldsymbol{x}_i \,-\, y_i\, e^{-\boldsymbol{x_i'\beta}}\, \boldsymbol{x_i}\right), \quad \frac{\partial^2 G}{\partial \boldsymbol{\beta}^2} = -\sum_{i=1}^{n}\left(y_i\, e^{-\boldsymbol{x_i'\beta}}\, \boldsymbol{x_i x_i'}\right).
$$

Let us approximate $e^{-\boldsymbol{x_i'\beta}} \approx (1 - \boldsymbol{x_i'\beta})$, the first-order Taylor's series approximation at $\boldsymbol{\beta} = \mathbf{0}$. Then the approximate maximum likelihood estimation (MLE) of $\boldsymbol{\beta}$ is

$$
\boldsymbol{\beta}^* = \left[\sum_{i=1}^{n} y_i(\boldsymbol{x_i x_i'})\right]^{-1} \left(\sum_{i=1}^{n}(y_i - 1)\boldsymbol{x_i}\right). \qquad (2.4)
$$

Let us denote the gradient vector $\boldsymbol{g_\beta} = \frac{\partial G}{\partial \boldsymbol{\beta}}|_{\boldsymbol{\beta}=\boldsymbol{\beta^*}}$, and the Hessian matrix $H(\boldsymbol{\beta}^*) = \frac{\partial^2 G}{\partial \boldsymbol{\beta}^2}|_{\boldsymbol{\beta}=\boldsymbol{\beta^*}}$. For simplicity let us denote $H(\boldsymbol{\beta}^*)$ by H^*. Using the *multivariate normal approximation theorem* from Chapter 1, we approximate the posterior density of $\boldsymbol{\beta}$ as

$$
\pi(\boldsymbol{\beta}|\boldsymbol{y}) \propto e^{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\mu_\beta})'(-H^*)(\boldsymbol{\beta}-\boldsymbol{\mu_\beta})}, \qquad (2.5)
$$

where $\boldsymbol{\mu_\beta} = \boldsymbol{\beta}^* + (-H^*)^{-1}\boldsymbol{g_\beta}$. From this approximated posterior density function, we can derive the approximated multivariate normal distribution for parameter $\boldsymbol{\beta}$ as

$$
\boldsymbol{\beta} \sim N(\boldsymbol{\mu_\beta},\ -H^{*-1}). \qquad (2.6)
$$

2.1.1 Sampling from Joint Posterior Density

We have only the vector $\boldsymbol{\beta}$ in this model. We draw $\boldsymbol{\beta}$ using the Metropolis–Hastings algorithm. The proposal distribution for $\boldsymbol{\beta}$ is the multivariate t-distribution with d degrees of freedom, with the mean and covariance matrix as in (2.6). The target density is the posterior density function (2.3). We draw a total of 1,000 samples, and we keep samples only if they move in the Markov chain Monte Carlo (MCMC) sequence. We check the acceptance rate of the MH algorithm and test the convergence of the MCMC sequence.

2.1.2 Prediction

After drawing a set of parameters from the hierarchical Bayesian exponential model, we predict response variables as follows:

(i) Find the rate parameters, $\lambda_i = e^{-\boldsymbol{x}_i'\boldsymbol{\beta}}$.

(ii) Predict responses from the exponential distribution, $\hat{y}_i \sim \text{Expo}(\lambda_i)$.

2.2 Exponential Model with Random Area Effects

The exponential model with random area effects is the simplest model we have with random area effects. We assume the responses $y_{ij}|\boldsymbol{\beta}, \nu_i$ are random samples from the exponential distribution with rate $e^{-(\boldsymbol{x}_{ij}'\boldsymbol{\beta}+\nu_i)}$. We assume that the random area effect ν_i follow the normal distribution with mean zero and variance σ^2. The likelihood function is given by

$$f(\boldsymbol{y}|\boldsymbol{\beta},\boldsymbol{\nu}) = \prod_{i=1}^{\ell}\prod_{j=1}^{n_i}\left[e^{-\boldsymbol{x}_{ij}'\boldsymbol{\beta}}\, e^{-(\nu_i + e^{-\nu_i}\, y_{ij}\, e^{-\boldsymbol{x}_{ij}'\boldsymbol{\beta}})}\right]. \tag{2.7}$$

We assume $\boldsymbol{\beta}$ and σ^2 have non-informative independent priors. The hierarchical Bayesian exponential model with random area effects is

$$\begin{aligned}
y_{ij}|\boldsymbol{\beta}, \nu_i &\overset{\text{ind}}{\sim} \text{Expo}\left(e^{-(\boldsymbol{x}_{ij}'\boldsymbol{\beta}+\boldsymbol{\nu_i})}\right), \quad \lambda_{ij} = e^{-(\boldsymbol{x}_{ij}'\boldsymbol{\beta}+\boldsymbol{\nu_i})},\ i = 1, \cdots, \ell,\ j = 1, \cdots, n_i, \\
\nu_i &\overset{\text{iid}}{\sim} N(0, \sigma^2), \\
\pi(\boldsymbol{\beta}, \sigma^2) &\propto \frac{1}{(1+\sigma^2)^2}.
\end{aligned} \tag{2.8}$$

Combining the likelihood in (2.7) and the priors in (2.8) via Bayes' theorem, we get the joint posterior density of $\boldsymbol{\beta}, \boldsymbol{\nu}$ and σ^2 given sample data as

$$\begin{aligned}
&\pi(\boldsymbol{\beta},\boldsymbol{\nu},\sigma^2|\boldsymbol{y}) \propto f(\boldsymbol{y}|\boldsymbol{\beta},\boldsymbol{\nu})\,\pi(\boldsymbol{\beta},\sigma^2)\,\pi(\boldsymbol{\nu}) \\
&\quad = \left[e^{-\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\boldsymbol{x'_{ij}\beta}}\right] \times \left[e^{-\sum_{i=1}^{\ell}(n_i\nu_i + e^{-\nu_i}\sum_{j=1}^{n_i} y_{ij}\, e^{-\boldsymbol{x'_{ij}\beta}})}\right] \times \prod_{i=1}^{\ell}\left[\left(\tfrac{1}{\sigma^2}\right)^{\frac{1}{2}} e^{-\frac{\nu_i^2}{2\sigma^2}}\right] \times \tfrac{1}{(1+\sigma^2)^2} \qquad (2.9)\\
&\quad = \tfrac{1}{(1+\sigma^2)^2}\left(\tfrac{1}{\sigma^2}\right)^{\frac{\ell}{2}} e^{-\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\boldsymbol{x'_{ij}\beta}}\; e^{-\sum_{i=1}^{\ell}(n_i\nu_i + e^{-\nu_i}\sum_{j=1}^{n_i} y_{ij}\, e^{-\boldsymbol{x'_{ij}\beta}} + \frac{\nu_i^2}{2\sigma^2})} \qquad (2.10)
\end{aligned}$$

Let the log-likelihood function be $G(\boldsymbol{\tau}|\boldsymbol{y}) = log(f(\boldsymbol{y}|\boldsymbol{\tau}))$, where $\boldsymbol{\tau} = (\boldsymbol{\beta'},\, \boldsymbol{\nu'})'$

$$G(\boldsymbol{\tau}|\boldsymbol{y}) = -\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\boldsymbol{x'_{ij}\beta} - \sum_{i=1}^{\ell}(n_i\nu_i + e^{-\nu_i}\sum_{j=1}^{n_i} y_{ij}\, e^{-\boldsymbol{x'_{ij}\beta}}).$$

For notational simplicity, let us write G for the log-likelihood function. Then its first- and second-order partial derivatives with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\nu}$ are

$$\begin{aligned}
\frac{\partial G}{\partial \boldsymbol{\beta}} &= -\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\boldsymbol{x_{ij}} + \sum_{i=1}^{\ell} e^{-\nu_i}\sum_{j=1}^{n_i} y_{ij}\, e^{-\boldsymbol{x'_{ij}\beta}}\,\boldsymbol{x_{ij}},\\
\frac{\partial G}{\partial \nu_i} &= -\left(n_i - e^{-\nu_i}\sum_{j=1}^{n_i} y_{ij}\, e^{-\boldsymbol{x'_{ij}\beta}}\right),\\
\frac{\partial^2 G}{\partial \boldsymbol{\beta}^2} &= -\sum_{i=1}^{\ell} e^{-\nu_i}\sum_{j=1}^{n_i} y_{ij}\, e^{-\boldsymbol{x'_{ij}\beta}}\,\boldsymbol{x_{ij}x'_{ij}},\\
\frac{\partial^2 G}{\partial \nu_i^2} &= -\,e^{-\nu_i}\sum_{j=1}^{n_i} y_{ij}\, e^{-\boldsymbol{x'_{ij}\beta}}, \text{ and}\\
\frac{\partial^2 G}{\partial \boldsymbol{\beta}\partial \nu_i} &= -\,e^{-\nu_i}\sum_{j=1}^{n_i} y_{ij}\, e^{-\boldsymbol{x'_{ij}\beta}}\,\boldsymbol{x_{ij}}.
\end{aligned}$$

Let us approximate $e^{-\boldsymbol{x'_{ij}\beta}} \approx 1 - \boldsymbol{x'_{ij}\beta}$, the first-order Taylor's series approximation at $\boldsymbol{\beta} = \boldsymbol{0}$, then the approximated MLE of $\boldsymbol{\beta}$ is

$$\boldsymbol{\beta}^* = \left[\sum_{i=1}^{\ell}\sum_{j=1}^{n_i} e^{-\nu_i} y_{ij}(\boldsymbol{x_i x'_i})\right]^{-1}\left(\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}(e^{-\nu_i}y_{ij} - 1)\boldsymbol{x_{ij}}\right). \qquad (2.11)$$

The MLE of ν_i is given by

$$\boldsymbol{\nu_i^*} = -\log\left[\frac{n_i}{\sum_{j=1}^{n_i} y_{ij}\, e^{-\boldsymbol{x'_{ij}\beta}}}\right]. \qquad (2.12)$$

Let the gradient vector be $\nabla G(\tau^*) = (\boldsymbol{g}_{\boldsymbol{\nu}}', \, \boldsymbol{g}_{\boldsymbol{\beta}}')'$. where $\boldsymbol{g}_{\boldsymbol{\nu}} = \left(\frac{\partial G}{\partial \nu_1} \;\; \cdots \;\; \frac{\partial G}{\partial \nu_\ell} \right)'|_{\boldsymbol{\nu}=\boldsymbol{\nu}^*, \, \boldsymbol{\beta}=\boldsymbol{\beta}^*}$, and $\boldsymbol{g}_{\boldsymbol{\beta}} = \left(\frac{\partial G}{\partial \beta_0} \;\; \cdots \;\; \frac{\partial G}{\partial \beta_p} \right)'|_{\boldsymbol{\nu}=\boldsymbol{\nu}^*, \, \boldsymbol{\beta}=\boldsymbol{\beta}^*}$, and the Hessian matrix $H(\tau^*)$ are evaluated at the approximate mode values $\boldsymbol{\beta}^*$ and $\boldsymbol{\nu}^*$. Then using the second-order Taylor's series approximation, we can write the approximated likelihood function as

$$\begin{aligned} f(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\nu}) &\approx e^{\left[G(\boldsymbol{\tau}^*) + \frac{1}{2}(\nabla G(\tau^*))' \, (-H(\boldsymbol{\tau}^*))^{-1} \, \nabla G(\tau^*)\right]} \\ &\qquad \times (2\pi)^{\frac{p+\ell}{2}} \, \left|(-H(\tau^*))^{-1}\right|^{\frac{1}{2}} \times N\left[\tau^* + (-H(\boldsymbol{\tau}^*))^{-1} \, \nabla G(\tau^*), \, (-H(\boldsymbol{\tau}^*))^{-1}\right], \end{aligned}$$

where N denotes the multivariate normal distribution for the parameter set $\boldsymbol{\tau} = (\boldsymbol{\beta}', \, \boldsymbol{\nu}')'$. Following the *multivariate normal approximation theorem* in Chapter 1 we can write

$$\begin{pmatrix} \boldsymbol{\nu} \\ \boldsymbol{\beta} \end{pmatrix} \sim N\left\{ \begin{pmatrix} \boldsymbol{\mu}_{\boldsymbol{\nu}}^* \\ \boldsymbol{\mu}_{\boldsymbol{\beta}}^* \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{pmatrix} \right\},$$

where the Hessian matrix is $H = -\begin{pmatrix} A_{11} & A_{12} \\ A_{12}' & A_{22} \end{pmatrix}$. Using the same notation as in Chapter 1, equations 1.3 and 1.4 for vectors and matrices, then applying the *multivariate normal approximation theorem*, we can write the approximated joint posterior density as

$$\begin{aligned} & f(\boldsymbol{\beta}, \boldsymbol{\nu}, \sigma^2|\boldsymbol{y}) \\ & \propto N\left(\boldsymbol{\mu}_\beta^*, \, \Sigma_{22}\right) \times N\left(\boldsymbol{\mu}_\nu^* + \Sigma_{12}\Sigma_{22}^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu}_\beta^*), \, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}'\right) \times N\left(\mathbf{0}, \, \sigma^2 I_\ell\right) \times \frac{1}{(1+\sigma^2)^2} \\ & \propto \frac{e^{-\frac{1}{2}\left[(\boldsymbol{\beta}-\boldsymbol{\mu}_\beta^*)' \, \Sigma_{22}^{-1} \, (\boldsymbol{\beta}-\boldsymbol{\mu}_\beta^*)\right]}}{|\Sigma_{22}|^{\frac{1}{2}} \left|A_{11}^{-1}\right|^{\frac{1}{2}}} \, e^{-\frac{1}{2}\left[(\boldsymbol{\nu}-(\boldsymbol{\mu}_\nu^*-A_{11}^{-1}A_{12}(\boldsymbol{\beta}-\boldsymbol{\mu}_\beta^*))' \, A_{11}^{-1} \, (\boldsymbol{\nu}-(\boldsymbol{\mu}_\nu^*-A_{11}^{-1}A_{12}(\boldsymbol{\beta}-\boldsymbol{\mu}_\beta^*))\right]} \\ & \quad \times \frac{e^{-\frac{1}{2}\left[\boldsymbol{\nu}'\left(\sigma^2 I_\ell\right)^{-1}\boldsymbol{\nu}\right]}}{|\sigma^2 I_\ell|^{\frac{1}{2}} \, (1+\sigma^2)^2} \\ & = \frac{|A_{11}|^{\frac{1}{2}}}{|\Sigma_{22}|^{\frac{1}{2}} \, |\sigma^2 I_\ell|^{\frac{1}{2}} \, (1+\sigma^2)^2} \times e^{-\frac{1}{2}\left[(\boldsymbol{\beta}-\boldsymbol{\mu}_\beta^*)' \, \Sigma_{22}^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu}_\beta^*)\right]} \\ & \quad \times e^{-\frac{1}{2}\left[(\boldsymbol{\mu}_\nu^*-A_{11}^{-1}A_{12}(\boldsymbol{\beta}-\boldsymbol{\mu}_\beta^*))'A_{11}\left(\left(A_{11}+\left(\sigma^2 I_\ell\right)^{-1}\right)\right)^{-1}\left(\sigma^2 I_\ell\right)^{-1}(\boldsymbol{\mu}_\nu^*-A_{11}^{-1}A_{12}(\boldsymbol{\beta}-\boldsymbol{\mu}_\beta^*))\right]} \\ & \quad \times e^{-\frac{1}{2}\left[\left[\boldsymbol{\nu}-\left(A_{11}+(\sigma^2 I_\ell)^{-1}\right)^{-1}(A_{11}\boldsymbol{\mu}_\nu^*-A_{12}(\boldsymbol{\beta}-\boldsymbol{\mu}_\beta^*))\right]'\left(A_{11}+(\sigma^2 I_\ell)^{-1}\right)\left[\boldsymbol{\nu}-\left(A_{11}+(\sigma^2 I_\ell)^{-1}\right)^{-1}(A_{11}\boldsymbol{\mu}_\nu^*-A_{12}(\boldsymbol{\beta}-\boldsymbol{\mu}_\beta^*))\right]\right]}. \end{aligned} \tag{2.13}$$

From the above joint posterior density, we notice that $\boldsymbol{\nu}$ has the multivariate normal distribution

$$\boldsymbol{\nu}|\boldsymbol{\beta}, \sigma^2 \sim N\left[\left(A_{11} + (\sigma^2 I_\ell)^{-1}\right)^{-1}(A_{11}\boldsymbol{\mu}_\nu^* - A_{12}(\boldsymbol{\beta} - \boldsymbol{\mu}_\beta^*)), \, \left(A_{11} + (\sigma^2 I_\ell)^{-1}\right)^{-1}\right]. \tag{2.14}$$

There are numerous small areas in our model. So we integrate out $\boldsymbol{\nu}$ from the joint posterior density (2.13), which will reduce the large number of parameters in the MCMC sequence. After integrating out $\boldsymbol{\nu}$, we have the joint distribution $\boldsymbol{\beta}, \sigma^2|\boldsymbol{y}$ given by

$$f(\boldsymbol{\beta}, \sigma^2|\boldsymbol{y}) \propto \frac{|A_{11}|^{\frac{1}{2}}}{|\Sigma_{22}|^{\frac{1}{2}}\ |\sigma^2 I_\ell|^{\frac{1}{2}}\ (1+\sigma^2)^2} \times e^{-\frac{1}{2}\left[(\boldsymbol{\beta}-\boldsymbol{\mu}^*_\beta)'\Sigma_{22}^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu}^*_\beta)\right]} \times \left|A_{11}+(\sigma^2 I_\ell)^{-1}\right|^{-\frac{1}{2}}$$
$$\times\, e^{-\frac{1}{2}\left[(\boldsymbol{\beta}-\widetilde{\boldsymbol{\mu}}_\beta)'\widetilde{\Sigma}(\boldsymbol{\beta}-\widetilde{\boldsymbol{\mu}}_\beta)\ -\ \widetilde{\boldsymbol{\mu}}'_\beta\widetilde{\Sigma}\widetilde{\boldsymbol{\mu}}_\beta\ +\ (\boldsymbol{\mu}^*_\nu+A_{11}^{-1}A_{12}\boldsymbol{\mu}^*_\beta)'S(\boldsymbol{\mu}^*_\nu+A_{11}^{-1}A_{12}\boldsymbol{\mu}^*_\beta)\right]},$$

where
$$S = A_{11}\left(A_{11}+(\sigma^2 I_\ell)^{-1}\right)^{-1}(\sigma^2 I_\ell)^{-1},$$
$$\widetilde{\boldsymbol{\mu}}_\beta = (A'_{12}A_{11}^{-1}SA_{11}^{-1}A_{12})^{-1}A'_{12}A_{11}^{-1}S\boldsymbol{\mu}^*_\nu + \boldsymbol{\mu}^*_\beta,$$
$$\widetilde{\Sigma}_\beta = A'_{12}A_{11}^{-1}SA_{11}^{-1}A_{12}.$$

From the above joint distribution of $\boldsymbol{\beta}, \sigma^2|\boldsymbol{y}$, we see that $\boldsymbol{\beta}$ has the multivariate normal distribution given by

$$\boldsymbol{\beta}|\sigma^2, \boldsymbol{y} \sim N\left[\left(\Sigma_{22}^{-1}+\widetilde{\Sigma}_\beta\right)^{-1}\left(\Sigma_{22}^{-1}\boldsymbol{\mu}^*_\beta+\widetilde{\Sigma}_\beta\widetilde{\boldsymbol{\mu}}_\beta\right),\ \left(\Sigma_{22}^{-1}+\widetilde{\Sigma}_\beta\right)^{-1}\right]. \tag{2.15}$$

Integrating out $\boldsymbol{\beta}$ from the above joint distribution of $\boldsymbol{\beta}, \sigma^2|\boldsymbol{y}$, we get the marginal distribution of $\sigma^2|\boldsymbol{y}$ as

$$\pi(\sigma^2|\boldsymbol{y})$$
$$\propto \frac{|A_{11}+(\sigma^2 I_\ell)^{-1}|^{-\frac{1}{2}}\left|\Sigma_{22}^{-1}+\widetilde{\Sigma}_\beta\right|^{-\frac{1}{2}}}{|\sigma^2 I_\ell|^{\frac{1}{2}}\ (1+\sigma^2)^2} \times e^{-\frac{1}{2}\left[(\boldsymbol{\mu}^*_\beta-\widetilde{\boldsymbol{\mu}}_\beta)'\Sigma_{22}^{-1}\left(\Sigma_{22}^{-1}+\widetilde{\Sigma}_\beta\right)^{-1}\widetilde{\Sigma}_\beta(\boldsymbol{\mu}^*_\beta-\widetilde{\boldsymbol{\mu}}_\beta)\right]}$$
$$\times\, e^{-\frac{1}{2}\left[-\widetilde{\boldsymbol{\mu}}'_\beta\widetilde{\Sigma}_\beta\widetilde{\boldsymbol{\mu}}_\beta+(\boldsymbol{\mu}^*_\nu+A_{11}^{-1}A_{12}\boldsymbol{\mu}^*_\beta)'S(\boldsymbol{\mu}^*_\nu+A_{11}^{-1}A_{12}\boldsymbol{\mu}^*_\beta)\right]}. \tag{2.16}$$

2.2.1 Sampling from Joint Posterior Density

We can draw approximated $\boldsymbol{\beta}$ and $\boldsymbol{\nu}$ samples from a multivariate normal distribution. However, $\sigma^2|\boldsymbol{y}$ is not in closed form. We use the grid sampling method and the Metropolis–Hastings sampling method to draw samples.

(i) We draw $\sigma^2|\boldsymbol{y}$ using the grid method. Since $\sigma^2 \in (0,\infty)$, we transform σ^2 into η which has range $(0,1)$, using the relation $\eta = \frac{\sigma^2}{1+\sigma^2}$. We take 1,000 grids of η and compute the transformed probability $\pi(\eta|\boldsymbol{y})$ using (2.16). We draw 1,000 samples of η with replacement from this grid probability distribution, then transform it back to σ^2.

(ii) Using the information about $\sigma^2|\boldsymbol{y}$ we can draw $\boldsymbol{\beta}|\sigma^2, \boldsymbol{y}$. The Metropolis–Hastings sampling method is used to draw jointly $\boldsymbol{\beta}, \sigma^2|\boldsymbol{y}$. The proposal densities are t-distributions. We take the log-transformation for the proposal density of σ^2. Then consider $\log(\sigma^2)|\boldsymbol{y}$ has a univariate t-distribution with d degrees of freedom, $\log(\sigma^2) \sim t_d(\mu_{ln}, \sigma^2_{ln})$, where μ_{ln} and σ^2_{ln} are estimated from samples of σ^2 in the above step. The proposal distribution for $\boldsymbol{\beta}|\sigma^2, \boldsymbol{y}$ is a multivariate t-distribution with d degrees of freedom, with the corresponding mean and covariance matrix as in equation (2.15). The target density is

$$\pi(\boldsymbol{\beta}, \sigma^2|\boldsymbol{y}) \propto \frac{e^{-\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\boldsymbol{x}'_{ij}\boldsymbol{\beta}}}{(1+\sigma^2)^2} \prod_{i=1}^{\ell}\left[\int_{\nu_i} e^{-(n_i\nu_i + e^{-\nu_i}\sum_{j=1}^{n_i} y_{ij}\, e^{-\boldsymbol{x}'_{ij}\boldsymbol{\beta}} + \frac{\nu_i^2}{2\sigma^2})} \times \left(\frac{1}{\sigma^2}\right)^{\frac{1}{2}} d\nu_i\right].$$

This integration is not in simple form. So we divide the integration domain into m equal intervals $[t_k,\, t_{k-1}]$ and apply the numerical integration

$$\pi(\boldsymbol{\beta}, \sigma^2|\boldsymbol{y}) \propto \frac{e^{-\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\boldsymbol{x}'_{ij}\boldsymbol{\beta}}}{(1+\sigma^2)^2} \prod_{i=1}^{\ell}\left[\sum_{k=1}^{m}\int_{t_{k-1}}^{t_k} e^{-(n_i\nu_i + e^{-\nu_i}\sum_{j=1}^{n_i} y_{ij}\, e^{-\boldsymbol{x}'_{ij}\boldsymbol{\beta}})} \times \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\nu_i^2}{2\sigma^2}} d\nu_i\right].$$

Since ν_i have the independent normal distributions, we can transform ν_i to the standard normal distribution, $z_i = \frac{\nu_i}{\sigma}$. For numerical integration we take the middle point of each interval $\hat{z}_k = \frac{t_{k-1}+t_k}{2}$. It gives

$$\begin{aligned}\pi(\boldsymbol{\beta}, \sigma^2|\boldsymbol{y}) &\propto \frac{e^{-\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\boldsymbol{x}'_{ij}\boldsymbol{\beta}}}{(1+\sigma^2)^2} \prod_{i=1}^{\ell}\left[\sum_{k=1}^{m} e^{-(n_i\hat{z}_k\sigma + e^{-\hat{z}_k\sigma}\sum_{j=1}^{n_i} y_{ij}\, e^{-\boldsymbol{x}'_{ij}\boldsymbol{\beta}})} \times \int_{t_{k-1}}^{t_k} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz\right] \\ &= \frac{e^{-\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\boldsymbol{x}'_{ij}\boldsymbol{\beta}}}{(1+\sigma^2)^2} \prod_{i=1}^{\ell}\left[\sum_{k=1}^{m} e^{-(n_i\hat{z}_k\sigma + e^{-\hat{z}_k\sigma}\sum_{j=1}^{n_i} y_{ij}\, e^{-\boldsymbol{x}'_{ij}\boldsymbol{\beta}})} \times (\Phi(t_k) - \Phi(t_{k-1}))\right].\end{aligned}$$

In the MH algorithm MCMC sequence, we keep the new sample only when it moves, and we check the acceptance rate of the MH sampler and test the convergence of the MCMC sequence.

(iii) Parameters $\nu_i|\boldsymbol{\beta}, \sigma^2$ are drawn using the Metropolis–Hastings algorithm method independently. The proposal density is a t-distribution with d degrees of freedom, with mean and variance of ν_i taken from the samples of the above step while drawing jointly $\boldsymbol{\beta}$ and σ^2. The target density is

$$\pi(\nu_i|\boldsymbol{\beta},\sigma^2) \propto e^{-\left(n_i\nu_i+e^{-\nu_i}\sum_{j=1}^{n_i} y_{ij}\, e^{-\boldsymbol{x}'_{ij}\boldsymbol{\beta}}+\frac{\nu_i^2}{2\sigma^2}\right)}, \quad i=1,\cdots,\ell.$$

We draw ν_i samples from the MH algorithm. If the acceptance rate of the sample drawing is not between 0.25 to 0.50, we discard the sample from the MH algorithm and draw ν_i using the grid sampling method at the second attempt.

2.2.2 Prediction

After drawing a set of samples as defined above from the hierarchical Bayesian exponential model with random area effects, we predict the responses as follows:

(i) Draw the random area effect. For each PSU we have one random area effects parameter, $\nu_i, i=1,\cdots,\ell$. We have sampled these parameters for all PSUs of NLSS-II. We will use these random area effects parameters to find the rate parameter.

(ii) Obtain the rate parameters directly as follows

$$\lambda_{ij} = e^{-(\boldsymbol{x}'_i\boldsymbol{\beta}+\nu_i)}.$$

(iii) Predict the responses from the exponential distribution

$$\hat{y}_{ij} \sim \text{Expo}(\lambda_{ij}).$$

2.3 Gamma Model without Random Area Effects

The gamma distribution is a generalization of the exponential distribution. It has two parameters, the shape α and the rate λ. It yields the exponential distribution when α equals one. We assume the responses $y_i|\boldsymbol{\beta},\alpha,\ i=1,\cdots,n$ are random samples from the gamma distribution with shape α and rate $e^{-\boldsymbol{x_i'\beta}}$. The likelihood function is

$$\begin{aligned} f(\boldsymbol{y}|\boldsymbol{\beta},\alpha) &= \prod_{i=1}^{n} \frac{e^{-y_i\, e^{-\boldsymbol{x_i'\beta}}}\, y_i^{\alpha-1}}{\Gamma(\alpha)}\, e^{-\alpha \boldsymbol{x_i'\beta}} \\ &= \left(\frac{g^{\alpha-1}}{\Gamma(\alpha)}\right)^n e^{-\sum_{i=1}^n \left(\alpha \boldsymbol{x_i'\beta} + y_i\, e^{-\boldsymbol{x_i'\beta}}\right)}, \end{aligned} \tag{2.17}$$

where $g = \left(\prod_{i=1}^n y_i\right)^{\frac{1}{n}}$ is the geometric mean of the response variable. We assume non-informative independent priors for $\boldsymbol{\beta}$ and α. The hierarchical Bayesian gamma model without random area effects is

$$\begin{aligned} y_i|\boldsymbol{\beta},\alpha &\overset{\text{ind}}{\sim} \text{Gamma}(\alpha, e^{-\boldsymbol{x_i'\beta}}), \quad \lambda_i = e^{-\boldsymbol{x_i'\beta}},\ i = 1,\cdots,n, \\ \pi(\boldsymbol{\beta},\alpha) &\propto \frac{1}{(1+\alpha)^2}. \end{aligned} \tag{2.18}$$

Combining the likelihood in (2.17) and the priors in (2.18) via Bayes' theorem, we get the joint posterior density of α and $\boldsymbol{\beta}$ given sample data as

$$\begin{aligned} \pi(\boldsymbol{\beta},\alpha|\boldsymbol{y}) &\propto f(\boldsymbol{y}|\boldsymbol{\beta},\alpha)\,\pi(\boldsymbol{\beta},\alpha) \\ &= \frac{1}{(1+\alpha)^2}\left(\frac{g^{\alpha-1}}{\Gamma(\alpha)}\right)^n e^{-\sum_{i=1}^n \left(\alpha \boldsymbol{x_i'\beta} + y_i\, e^{-\boldsymbol{x_i'\beta}}\right)}. \end{aligned} \tag{2.19}$$

Letting the log-likelihood function be $G(\alpha,\boldsymbol{\beta}|\boldsymbol{y}) = log(f(\boldsymbol{y}|\boldsymbol{\alpha},\boldsymbol{\beta}))$,

$$G(\alpha,\boldsymbol{\beta}|\boldsymbol{y}) = -\sum_{i=1}^n \left(\alpha \boldsymbol{x_i'\beta} + y_i\, e^{-\boldsymbol{x_i'\beta}}\right) + n\left[(\alpha-1)\log(g) - \log(\Gamma(\alpha))\right].$$

For notational simplicity, let us write G for the log-likelihood function. Then its first- and second-order partial derivatives with respect to $\boldsymbol{\beta}$ are given by

$$\frac{\partial G}{\partial \boldsymbol{\beta}} = -\sum_{i=1}^n \left(\alpha - y_i\, e^{-\boldsymbol{x_i'\beta}}\right)\boldsymbol{x_i}, \quad \frac{\partial^2 G}{\partial \boldsymbol{\beta}^2} = -\sum_{i=1}^n \left(y_i\, e^{-\boldsymbol{x_i'\beta}}\,\boldsymbol{x_i x_i'}\right).$$

Let us approximate $e^{-\boldsymbol{x_i'\beta}} \approx (1 - \boldsymbol{x_i'\beta})$, the first-order Taylor's series approximation at $\boldsymbol{\beta} = \mathbf{0}$. Then we get the approximated MLE of $\boldsymbol{\beta}|\alpha$

$$\boldsymbol{\beta}^*|\alpha = \left[\sum_{i=1}^n y_i(\boldsymbol{x_i x_i'})\right]^{-1}\left(\sum_{i=1}^n (y_i - \alpha)\boldsymbol{x_i}\right). \tag{2.20}$$

Let the gradient vector and the Hessian matrix be evaluated at the approximate mode values $\boldsymbol{\beta}^*$ be $\nabla G(\beta^*)$ and $H(\boldsymbol{\beta}^*)$ respectively. Using the *multivariate normal approximation* *theorem*, we can write the approximated likelihood function as

$$f(\boldsymbol{\beta}, \boldsymbol{\alpha}|\boldsymbol{y}) \propto \frac{1}{(1+\alpha)^2} \, e^{\left[G(\boldsymbol{\beta}^*)+\frac{1}{2}(\nabla G(\boldsymbol{\beta}^*))'\,(-H(\boldsymbol{\beta}^*))^{-1}\,\nabla G(\boldsymbol{\beta}^*)\right]} \\ \times \left|(-H(\boldsymbol{\beta}^*))^{-1}\right|^{\frac{1}{2}} \, N\left[\beta^* + (-H(\boldsymbol{\beta}^*))^{-1}\,\nabla G(\boldsymbol{\beta}^*),\, (-H(\boldsymbol{\beta}^*))^{-1}\right]. \tag{2.21}$$

From the above joint distribution it follows that $\boldsymbol{\beta}$ has the multivariate normal distribution given by

$$\boldsymbol{\beta}|\alpha, \boldsymbol{y} \sim \mathrm{N}\left[\beta^* + (-H(\boldsymbol{\beta}^*))^{-1}\,\nabla G(\boldsymbol{\beta}^*),\, (-H(\boldsymbol{\beta}^*))^{-1}\right], \tag{2.22}$$

where N denotes the multivariate normal distribution for parameter $\boldsymbol{\beta}$. Integrating out $\boldsymbol{\beta}$ we get the marginal distribution of $\alpha|\boldsymbol{y}$

$$f(\alpha|\boldsymbol{y}) \propto \frac{1}{(1+\alpha)^2} \, \left|(-H(\boldsymbol{\beta}^*))^{-1}\right|^{\frac{1}{2}} \, e^{\left[G(\boldsymbol{\beta}^*)+\frac{1}{2}(\nabla G(\boldsymbol{\beta}^*))'\,(-H(\boldsymbol{\beta}^*))^{-1}\,\nabla G(\boldsymbol{\beta}^*)\right]}.$$

2.3.1 Sampling from Joint Posterior Density

We use the grid method and the Metropolis–Hastings algorithm to draw samples from the joint posterior density function. The proposal density functions for the Metropolis–Hastings sampler are t-distributions. The steps for drawing samples are as follows:

(i) We draw $\alpha|\boldsymbol{y}$ using the grid method. Since $\alpha \in (0, \infty)$, we transform α into η which has range $(0, 1)$, $\eta = \frac{\alpha}{1+\alpha}$. We take 1,000 grids of η and compute transformed probability $\pi(\eta|\boldsymbol{y})$ using (2.23). We draw 1,000 samples of η from this grid probability distribution, then transform back to α.

(ii) Once we draw parameter α in the above step, then we draw $\boldsymbol{\beta}$ given α. We used the Metropolis–Hastings algorithm to sample α and $\boldsymbol{\beta}$ jointly. The proposal densities are t-distributions. We take the log-transformation for the proposal density of α. Then consider $\log(\alpha)|\boldsymbol{y}$ has a univariate t-distribution with d degrees of freedom $\log(\alpha) \sim t_d(\mu_{ln}, \sigma^2_{ln})$, where μ_{ln} and σ^2_{ln} are the estimated mean and variance from the above step. The proposal density for $\boldsymbol{\beta}|\boldsymbol{y}, \sigma^2$ is a multivariate t-distribution with d degrees of freedom. The corresponding mean and the covariance matrix are as in equation (2.22). The target density is the posterior distribution (2.19). In the MH algorithm, we accept the new sample only when it moves. We check the acceptance rate of the MH algorithm and test the convergence of the MCMC sequence.

2.3.2 Prediction

After drawing the set of parameters α and $\boldsymbol{\beta}$ from the gamma model without random area effects, we predict the responses as follows:

(i) Find the rate parameters, $\lambda_i = e^{-\boldsymbol{x}_i'\boldsymbol{\beta}}$.

(ii) Draw predicted responses from the gamma distribution $\hat{y}_i \sim \text{Gamma}(\alpha, \lambda_i)$.

2.4 Gamma Model with Random Area Effects

Here we develop the hierarchical Bayesian model with random area effects assuming the response variable has the gamma distribution. We assume that the responses $y_{ij}|\alpha, \lambda_{ij}$ are random samples from the gamma distribution with shape α and rate $e^{-(\boldsymbol{x}_{ij}'\boldsymbol{\beta}+\nu_i)}$ and assume that the random area effect ν_i follows the normal distribution with mean zero and variance σ^2. The likelihood function is given by

$$\begin{aligned} f(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\nu}, \alpha) &= \prod_{i=1}^{\ell} \prod_{j=1}^{n_i} \frac{e^{-y_{ij}\, e^{-(\boldsymbol{x}_{ij}'\boldsymbol{\beta}+\nu_i)}}\, y_{ij}^{\alpha-1}}{\Gamma(\alpha)}\, e^{-\alpha(\boldsymbol{x}_{ij}'\boldsymbol{\beta}+\nu_i)} \\ &= \left(\frac{g^{\alpha-1}}{\Gamma(\alpha)}\right)^n e^{-\alpha \sum_{i=1}^{\ell}\sum_{j=1}^{n_i} \boldsymbol{x}_{ij}'\boldsymbol{\beta}}\, e^{-\sum_{i=1}^{\ell}\left(\alpha n_i \nu_i + e^{-\nu_i}\sum_{j=1}^{n_i} y_{ij} e^{-\boldsymbol{x}_{ij}'\boldsymbol{\beta}}\right)}, \end{aligned} \tag{2.23}$$

where $g = \left(\prod_{i=1}^{\ell}\prod_{j=1}^{n_i} y_{ij}\right)^{\frac{1}{n}}$ is the geometric mean of the response variable. We assume the priors for $\alpha, \boldsymbol{\beta}$ and σ^2 are independent and non-informative. The hierarchical Bayesian gamma model with random area effects is

$$\begin{aligned} y_{ij}|\boldsymbol{\beta}, \alpha, \nu_i &\overset{\text{ind}}{\sim} \text{Gamma}\left(\alpha,\, e^{-(\boldsymbol{x}_{ij}'\boldsymbol{\beta}+\nu_i)}\right),\ \lambda_{ij} = e^{-(\boldsymbol{x}_{ij}'\boldsymbol{\beta}+\nu_i)},\ i = 1, \cdots, \ell,\ j = 1, \cdots, n_i, \\ \nu_i &\overset{\text{iid}}{\sim} N(0, \sigma^2), \\ \pi(\boldsymbol{\beta}, \alpha, \sigma^2) &\propto \frac{1}{(1+\sigma^2)^2(1+\alpha)^2}. \end{aligned} \tag{2.24}$$

Combining the likelihood in (2.23) and the priors in (2.24) via Bayes' theorem, we get the joint posterior density of $\alpha, \boldsymbol{\beta}, \boldsymbol{\nu}$ and σ^2 given sample data as

$$
\begin{aligned}
&\pi(\boldsymbol{\beta}, \boldsymbol{\nu}, \alpha|\boldsymbol{y}) \propto f(\boldsymbol{y}|\boldsymbol{\beta}, \alpha)\, \pi(\boldsymbol{\beta}, \alpha)\, \pi(\boldsymbol{\nu}) \\
&= \left(\tfrac{g^{\alpha-1}}{\Gamma(\alpha)}\right)^n e^{-\alpha \sum_{i=1}^{\ell}\sum_{j=1}^{n_i} \boldsymbol{x'_{ij}\beta}}\, e^{-\sum_{i=1}^{\ell}\left(\alpha n_i \nu_i + e^{-\nu_i}\sum_{j=1}^{n_i} y_{ij} e^{-\boldsymbol{x'_{ij}\beta}}\right)} \times \prod_{i=1}^{\ell}\left[\left(\tfrac{1}{\sigma^2}\right)^{\frac{1}{2}} e^{-\frac{\nu_i^2}{2\sigma^2}}\right] \tfrac{1}{(1+\sigma^2)^2\,(1+\alpha)^2} \\
&= \left(\tfrac{g^{\alpha-1}}{\Gamma(\alpha)}\right)^n e^{-\alpha \sum_{i=1}^{\ell}\sum_{j=1}^{n_i} \boldsymbol{x'_{ij}\beta}}\, e^{-\sum_{i=1}^{\ell}\left(\alpha n_i \nu_i + e^{-\nu_i}\sum_{j=1}^{n_i} y_{ij} e^{-\boldsymbol{x'_{ij}\beta}} + \frac{\nu_i^2}{2\sigma^2}\right)} \times \left(\tfrac{1}{\sigma^2}\right)^{\frac{\ell}{2}} \tfrac{1}{(1+\sigma^2)^2\,(1+\alpha)^2}. \qquad (2.25)
\end{aligned}
$$

Let the log-likelihood function be $G(\alpha, \boldsymbol{\tau}|\boldsymbol{y}) = log(f(\boldsymbol{y}|\boldsymbol{\alpha}, \boldsymbol{\tau}))$, where $\boldsymbol{\tau} = (\boldsymbol{\beta'},\, \boldsymbol{\nu'})'$

$$G(\alpha, \boldsymbol{\tau}|\boldsymbol{y}) = n\left[(\alpha-1)\log(g) - \log(\Gamma(\alpha))\right] - \alpha \sum_{i=1}^{\ell}\sum_{j=1}^{n_i} \boldsymbol{x'_{ij}\beta} - \sum_{i=1}^{\ell}\left(\alpha n_i \nu_i \,+\, e^{-\nu_i}\sum_{j=1}^{n_i} y_{ij} e^{-\boldsymbol{x'_{ij}\beta}}\right).$$

For notational simplicity, let us write G for the log-likelihood function. Then its first- and second-order partial derivatives with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\nu}$ are as follows:

$$
\begin{aligned}
\frac{\partial G}{\partial \boldsymbol{\beta}} &= -\,\alpha \sum_{i=1}^{\ell}\sum_{j=1}^{n_i} \boldsymbol{x_{ij}} + \sum_{i=1}^{\ell} e^{-\nu_i} \sum_{j=1}^{n_i} y_{ij}\, e^{-\boldsymbol{x'_{ij}\beta}}\, \boldsymbol{x_{ij}}, \\
\frac{\partial G}{\partial \nu_i} &= -\left(\alpha\, n_i - e^{-\nu_i} \sum_{j=1}^{n_i} y_{ij}\, e^{-\boldsymbol{x'_{ij}\beta}}\right), \\
\frac{\partial^2 G}{\partial \boldsymbol{\beta}^2} &= -\sum_{i=1}^{\ell} e^{-\nu_i} \sum_{j=1}^{n_i} y_{ij}\, e^{-\boldsymbol{x'_{ij}\beta}}\, \boldsymbol{x_{ij}x'_{ij}}, \\
\frac{\partial^2 G}{\partial \nu_i^2} &= -\, e^{-\nu_i} \sum_{j=1}^{n_i} y_{ij}\, e^{-\boldsymbol{x'_{ij}\beta}}, \\
\frac{\partial^2 G}{\partial \boldsymbol{\beta}\partial \nu_i} &= -\, e^{-\nu_i} \sum_{j=1}^{n_i} y_{ij}\, e^{-\boldsymbol{x'_{ij}\beta}}\, \boldsymbol{x_{ij}}.
\end{aligned}
$$

Let $e^{-\boldsymbol{x'_i\beta}} \approx (1 - \boldsymbol{x'_i\beta})$, the first-order Taylor's series approximation at $\boldsymbol{\beta} = \mathbf{0}$, then the approximated MLE of $\boldsymbol{\beta}|\alpha$ is

$$\boldsymbol{\beta}^*|\alpha = \left[\sum_{i=1}^{\ell}\sum_{j=1}^{n_i} e^{-\nu_i} y_{ij} (\boldsymbol{x_i x'_i})\right]^{-1} \left(\sum_{i=1}^{\ell}\sum_{j=1}^{n_i} (e^{-\nu_i} y_{ij} - \alpha)\boldsymbol{x_{ij}}\right). \qquad (2.26)$$

The approximated MLE of ν_i is given by

$$\boldsymbol{\nu_i^*} = -\log\left[\frac{\alpha\, n_i}{\sum_{j=1}^{n_i} y_{ij}\, e^{-\boldsymbol{x'_{ij}\beta}}}\right]. \qquad (2.27)$$

Let the gradient vector be $\nabla G(\tau^*) = (\boldsymbol{g}'_{\boldsymbol{\nu}},\, \boldsymbol{g}'_{\boldsymbol{\beta}})'$, where $\boldsymbol{g}_{\boldsymbol{\nu}} = \left(\frac{\partial G}{\partial \nu_1} \;\; \cdots \;\; \frac{\partial G}{\partial \nu_\ell} \right)'|_{\boldsymbol{\nu}=\boldsymbol{\nu}^*,\, \boldsymbol{\beta}=\boldsymbol{\beta}^*}$, and $\boldsymbol{g}_{\boldsymbol{\beta}} = \left(\frac{\partial G}{\partial \beta_0} \;\; \cdots \;\; \frac{\partial G}{\partial \beta_p} \right)'|_{\boldsymbol{\nu}=\boldsymbol{\nu}^*,\, \boldsymbol{\beta}=\boldsymbol{\beta}^*}$, and the Hessian matrix be $H(\boldsymbol{\tau}^*)$ evaluated at approximate mode values $\boldsymbol{\beta}^*$ and $\boldsymbol{\nu}^*$. Using the second-order Taylor's series approximation, we can write the approximated likelihood function as

$$
\begin{aligned}
f(\boldsymbol{y}|\boldsymbol{\beta},\boldsymbol{\nu}) &\approx e^{\left[G_\alpha(\boldsymbol{\tau^*}) + \frac{1}{2}(\nabla G_\alpha(\tau^*))'\,(-H_\alpha(\boldsymbol{\tau^*}))^{-1}\,\nabla G_\alpha(\tau^*)\right]} \\
&\qquad \times (2\pi)^{\frac{p+\ell}{2}} \left|(-H_\alpha(\tau^*))^{-1}\right|^{\frac{1}{2}} N\left[\tau^* + (-H_\alpha(\boldsymbol{\tau^*}))^{-1}\,\nabla G_\alpha(\tau^*),\, (-H_\alpha(\boldsymbol{\tau^*}))^{-1}\right],
\end{aligned}
$$

where N denotes the multivariate normal distribution for the parameter set $\boldsymbol{\tau} = (\boldsymbol{\beta'},\, \boldsymbol{\nu'})'$. Following the *multivariate normal approximation theorem* in Chapter 1, we can write

$$
\begin{pmatrix} \boldsymbol{\nu} \\ \boldsymbol{\beta} \end{pmatrix} \sim N\left\{ \begin{pmatrix} \boldsymbol{\mu}^*_{\boldsymbol{\nu}} \\ \boldsymbol{\mu}^*_{\boldsymbol{\beta}} \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma'_{12} & \Sigma_{22} \end{pmatrix} \right\},
$$

where the Hessian matrix is $H = -\begin{pmatrix} A_{11} & A_{12} \\ A'_{12} & A_{22} \end{pmatrix}$. Let

$$
C_\alpha(\tau^*) = e^{\left[G_\alpha(\boldsymbol{\tau^*}) + \frac{1}{2}(\nabla G_\alpha(\tau^*))'\,(-H_\alpha(\boldsymbol{\tau^*}))^{-1}\,\nabla G_\alpha(\tau^*)\right]} \left|(-H_\alpha(\tau^*))^{-1}\right|^{\frac{1}{2}}.
$$

Using the same notation as in Chapter 1, equations 1.3 and 1.4 for vectors and matrices then applying the *multivariate normal approximation theorem*, we can write the approximated joint posterior density as

$$
\begin{aligned}
&f(\boldsymbol{\beta},\boldsymbol{\nu},\alpha,\sigma^2|\boldsymbol{y}) \\
&\propto C_\alpha(\tau^*) \times N\left(\boldsymbol{\mu}^*_\beta,\ \Sigma_{22}\right) \times N\left(\boldsymbol{\mu}^*_\nu + \Sigma_{12}\Sigma_{22}^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu}^*_\beta),\ \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma'_{12}\right) \times N\left(\mathbf{0},\, \sigma^2 I_\ell\right) \\
&\quad \times \frac{1}{(1+\sigma^2)^2\,(1+\alpha)^2} \\
&= C_\alpha(\tau^*) \times \frac{1}{(1+\sigma^2)^2\,(1+\alpha)^2} \times \frac{|A_{11}|^{\frac{1}{2}}}{|\Sigma_{22}|^{\frac{1}{2}}\ |\sigma^2 I_\ell|^{\frac{1}{2}}} \times e^{-\frac{1}{2}\left[(\boldsymbol{\beta}-\boldsymbol{\mu}^*_\beta)'\,\Sigma_{22}^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu}^*_\beta)\right]} \\
&\quad \times e^{-\frac{1}{2}\left[(\boldsymbol{\mu}^*_\nu - A_{11}^{-1}A_{12}(\boldsymbol{\beta}-\boldsymbol{\mu}^*_\beta))' A_{11}\left(\left(A_{11}+(\sigma^2 I_\ell\right)^{-1}\right)^{-1}\left(\sigma^2 I_\ell\right)^{-1}(\boldsymbol{\mu}^*_\nu - A_{11}^{-1}A_{12}(\boldsymbol{\beta}-\boldsymbol{\mu}^*_\beta))\right]} \\
&\quad \times e^{-\frac{1}{2}\left[\left[\boldsymbol{\nu}-\left(A_{11}+(\sigma^2 I_\ell)^{-1}\right)^{-1}(A_{11}\boldsymbol{\mu}^*_\nu - A_{12}(\boldsymbol{\beta}-\boldsymbol{\mu}^*_\beta))\right]'\left(A_{11}+(\sigma^2 I_\ell)^{-1}\right)\left[\boldsymbol{\nu}-\left(A_{11}+(\sigma^2 I_\ell)^{-1}\right)^{-1}(A_{11}\boldsymbol{\mu}^*_\nu - A_{12}(\boldsymbol{\beta}-\boldsymbol{\mu}^*_\beta))\right]\right]}.
\end{aligned}
\tag{2.28}
$$

From the above joint posterior density, we see that $\boldsymbol{\nu}$ has multivariate normal distribution

$$
\boldsymbol{\nu}|\boldsymbol{\beta},\alpha,\sigma^2 \sim N\left[\left(A_{11} + (\sigma^2 I_\ell)^{-1}\right)^{-1}(A_{11}\boldsymbol{\mu}^*_\nu - A_{12}(\boldsymbol{\beta} - \boldsymbol{\mu}^*_\beta)),\ \left(A_{11} + (\sigma^2 I_\ell)^{-1}\right)^{-1}\right] \tag{2.29}
$$

As we mentioned before, there are numerous small areas. Integrating out $\boldsymbol{\nu}$, we have a joint density of $\boldsymbol{\beta}, \alpha, \sigma^2 | \boldsymbol{y}$ as

$$\begin{aligned}
& f(\boldsymbol{\beta}, \alpha, \sigma^2 | \boldsymbol{y}) \\
& \quad \propto C_\alpha(\tau^*) \times \frac{1}{(1+\sigma^2)^2\,(1+\alpha)^2} \times \frac{|A_{11}|^{\frac{1}{2}}}{|\Sigma_{22}|^{\frac{1}{2}}\ |\sigma^2 I_\ell|^{\frac{1}{2}}} \times e^{-\frac{1}{2}\left[(\boldsymbol{\beta}-\boldsymbol{\mu}_\beta^*)'\,\Sigma_{22}^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu}_\beta^*)\right]} \\
& \qquad \times \left|A_{11} + (\sigma^2 I_\ell)^{-1}\right|^{-\frac{1}{2}} \times e^{-\frac{1}{2}\left[(\boldsymbol{\beta}-\widetilde{\boldsymbol{\mu}}_\beta)'\widetilde{\Sigma}(\boldsymbol{\beta}-\widetilde{\boldsymbol{\mu}}_\beta)\ -\ \widetilde{\boldsymbol{\mu}}_\beta'\widetilde{\Sigma}\widetilde{\boldsymbol{\mu}}_\beta\ +\ (\boldsymbol{\mu}_\nu^*+A_{11}^{-1}A_{12}\boldsymbol{\mu}_\beta^*)'S(\boldsymbol{\mu}_\nu^*+A_{11}^{-1}A_{12}\boldsymbol{\mu}_\beta^*)\right]},
\end{aligned}$$

where

$$\begin{aligned}
S &= A_{11}\left(A_{11} + (\sigma^2 I_\ell)^{-1}\right)^{-1}(\sigma^2 I_\ell)^{-1}, \\
\widetilde{\boldsymbol{\mu}}_\beta &= (A_{12}'A_{11}^{-1}SA_{11}^{-1}A_{12})^{-1}A_{12}'A_{11}^{-1}S\boldsymbol{\mu}_\nu^* + \boldsymbol{\mu}_\beta^*, \\
\widetilde{\Sigma}_\beta &= A_{12}'A_{11}^{-1}SA_{11}^{-1}A_{12}.
\end{aligned}$$

From the above joint density of $\boldsymbol{\beta}, \alpha, \sigma^2$, we notice that $\boldsymbol{\beta}$ has a multivariate normal distribution as

$$\boldsymbol{\beta}|\alpha, \sigma^2, \boldsymbol{y} \sim N\left[\left(\Sigma_{22}^{-1} + \widetilde{\Sigma}_\beta\right)^{-1}\left(\Sigma_{22}^{-1}\boldsymbol{\mu}_\beta^* + \widetilde{\Sigma}_\beta\widetilde{\boldsymbol{\mu}}_\beta\right),\ \left(\Sigma_{22}^{-1} + \widetilde{\Sigma}_\beta\right)^{-1}\right]. \tag{2.30}$$

Integrating out $\boldsymbol{\beta}$ from above joint distribution, we get the joint distribution of $\alpha, \sigma^2|\boldsymbol{y}$

$$\begin{aligned}
& \pi(\alpha, \sigma^2 | \boldsymbol{y}) \\
& \quad \propto C_\alpha(\tau^*) \times \frac{|A_{11}|^{\frac{1}{2}}}{|\Sigma_{22}|^{\frac{1}{2}}\ |\sigma^2 I_\ell|^{\frac{1}{2}}} \times \frac{1}{(1+\sigma^2)^2}\frac{1}{(1+\alpha)^2} \times \left|A_{11} + (\sigma^2 I_\ell)^{-1}\right|^{-\frac{1}{2}}\left|\Sigma_{22}^{-1} + \widetilde{\Sigma}_\beta\right|^{-\frac{1}{2}} \\
& \qquad \times e^{-\frac{1}{2}\left[(\boldsymbol{\mu}_\beta^*-\widetilde{\boldsymbol{\mu}}_\beta)'\Sigma_{22}^{-1}\left(\Sigma_{22}^{-1}+\widetilde{\Sigma}_\beta\widetilde{\boldsymbol{\mu}}_\beta\right)^{-1}\widetilde{\Sigma}_\beta(\boldsymbol{\mu}_\beta^*-\widetilde{\boldsymbol{\mu}}_\beta)\right]} \\
& \qquad \times e^{-\frac{1}{2}\left[-\ \widetilde{\boldsymbol{\mu}}_\beta'\widetilde{\Sigma}_\beta\widetilde{\boldsymbol{\mu}}_\beta\ +\ (\boldsymbol{\mu}_\nu^*+A_{11}^{-1}A_{12}\boldsymbol{\mu}_\beta^*)'S(\boldsymbol{\mu}_\nu^*+A_{11}^{-1}A_{12}\boldsymbol{\mu}_\beta^*)\right]}.
\end{aligned} \tag{2.31}$$

2.4.1 Sampling from Joint Posterior Density

The joint density of α and σ^2 is complicated. We use the parameter α sampled in the previous model, the hierarchical Bayesian model without random area effects, as an approximation. We can draw approximate $\boldsymbol{\beta}$ and $\boldsymbol{\nu}$ samples from a multivariate normal distribution. However, the CPD of $\sigma^2|\alpha, \boldsymbol{y}$ is not in a closed form. We use the grid method and the Metropolis–Hastings algorithm to draw samples.

(i) Borrow α samples from the previous hierarchical Bayesian gamma model without random area effects as an approximation. In 1,000 samples drawn in that model, we pick up 100 quantile values as an approximate α samples for this model.

(ii) Draw σ^2 from the grid method for given α in the above step. Samples $\sigma^2|\alpha, \boldsymbol{y}$ are drawn using the grid sampling method. The CPD is

$$\pi(\sigma^2|\alpha, \boldsymbol{y})$$
$$\propto \frac{1}{(1+\sigma^2)^2} \frac{|A_{11} + (\sigma^2 I_\ell)^{-1}|^{-\frac{1}{2}} \left|\Sigma_{22}^{-1} + \widetilde{\Sigma}_\beta\right|^{-\frac{1}{2}}}{|\sigma^2 I_\ell|^{\frac{1}{2}}} \times e^{-\frac{1}{2}\left[(\boldsymbol{\mu}_\beta^* - \widetilde{\boldsymbol{\mu}}_\beta)' \Sigma_{22}^{-1} \left(\Sigma_{22}^{-1} + \widetilde{\Sigma}_\beta\right)^{-1} \widetilde{\Sigma}_\beta (\boldsymbol{\mu}_\beta^* - \widetilde{\boldsymbol{\mu}}_\beta)\right]}$$
$$\times e^{-\frac{1}{2}\left[-\widetilde{\boldsymbol{\mu}}_\beta' \widetilde{\Sigma}_\beta \widetilde{\boldsymbol{\mu}}_\beta + (\boldsymbol{\mu}_\nu^* + A_{11}^{-1} A_{12} \boldsymbol{\mu}_\beta^*)' S (\boldsymbol{\mu}_\nu^* + A_{11}^{-1} A_{12} \boldsymbol{\mu}_\beta^*)\right]}. \qquad (2.32)$$

Since $\sigma^2 \in (0, \infty)$, we transform σ^2 into η, which has range $(0, 1)$, using the relation $\eta = \frac{\sigma^2}{1+\sigma^2}$. We take 100 grids of η for each α quantile value and compute transformed probability $\pi(\eta|\alpha, \boldsymbol{y})$ using (2.32).We draw a η sample from this grid probability distribution then transform it back to σ^2.

(iii) Using the information $\alpha, \sigma^2|\boldsymbol{y}$ drawn above, we can draw $\boldsymbol{\beta}|\alpha, \sigma^2, \boldsymbol{y}$. The Metropolis–Hastings algorithm is then used to draw jointly $\boldsymbol{\beta}, \alpha, \sigma^2|\boldsymbol{y}$. The proposal densities are t-distributions. We take the log-transformation for the joint proposal density of α and σ^2. Then consider $\log(\sigma^2, \alpha)|\boldsymbol{y}$ has a bivariate t-distribution with d degrees of freedom, $[\log(\sigma^2), log(\alpha)]|\boldsymbol{y} \sim \mathrm{t}_d(\boldsymbol{\mu}_{ln}, \Sigma_{ln})$, where $\boldsymbol{\mu}_{ln}$ and Σ_{ln} are estimated from the above samples. The proposal distribution for $\boldsymbol{\beta}|\boldsymbol{y}, \alpha, \sigma^2$ is also a multivariate t-distribution with d degrees of freedom, with a corresponding mean and covariance matrix as in equation (2.30). The target density is

$$\pi(\boldsymbol{\beta}, \alpha, \sigma^2|\boldsymbol{y}) \propto \left(\frac{g^{\alpha-1}}{\Gamma(\alpha)}\right)^n \frac{e^{-\alpha \sum_{i=1}^{\ell} \sum_{j=1}^{n_i} \boldsymbol{x}_{ij}' \boldsymbol{\beta}}}{(1+\sigma^2)^2 (1+\alpha)^2}$$
$$\times \prod_{i=1}^{\ell} \left[\int_{\nu_i} e^{-(\alpha n_i \nu_i + e^{-\nu_i} \sum_{j=1}^{n_i} y_{ij} e^{-\boldsymbol{x}_{ij}' \boldsymbol{\beta}} + \frac{\nu_i^2}{2\sigma^2})} \times \left(\frac{1}{\sigma^2}\right)^{\frac{1}{2}} d\nu_i\right].$$

This integration is not in a simple form. We apply a numerical integration. We divide the integration domain into m equal intervals $[t_k,\, t_{k-1}]$

$$\pi(\boldsymbol{\beta}, \sigma^2|\boldsymbol{y}) \propto \left(\frac{g^{\alpha-1}}{\Gamma(\alpha)}\right)^n \frac{e^{-\sum_{i=1}^{\ell}\sum_{j=1}^{n_i} \boldsymbol{x'_{ij}\beta}}}{(1+\sigma^2)^2\,(1+\alpha)^2} \times \prod_{i=1}^{\ell}\left[\sum_{k=1}^{m}\int_{t_{k-1}}^{t_k} e^{-(\alpha\, n_i\nu_i + e^{-\nu_i}\sum_{j=1}^{n_i} y_{ij}\, e^{-\boldsymbol{x'_{ij}\beta}})} \times \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\nu_i^2}{2\sigma^2}} d\nu_i\right].$$

We have assumed that ν_i has a univariate normal distribution centered at zero. We transform ν_i to the standard normal distribution, $z_i = \frac{\nu_i}{\sigma}$. For numerical integration, we take the middle point of each interval, $\hat{z}_k = \frac{t_{k-1}+t_k}{2}$,

$$\begin{aligned}\pi(\boldsymbol{\beta}, \sigma^2|\boldsymbol{y}) &\propto \left(\frac{g^{\alpha-1}}{\Gamma(\alpha)}\right)^n \frac{e^{-\sum_{i=1}^{\ell}\sum_{j=1}^{n_i} \boldsymbol{x'_{ij}\beta}}}{(1+\sigma^2)^2\,(1+\alpha)^2} \\ &\quad\times \left[\sum_{k=1}^{m} e^{-(\alpha\, n_i\hat{z}_k\sigma + e^{-\hat{z}_k\sigma}\sum_{j=1}^{n_i} y_{ij}\, e^{-\boldsymbol{x'_{ij}\beta}})} \times \int_{t_{k-1}}^{t_k} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz\right] \\ &= \left(\frac{g^{\alpha-1}}{\Gamma(\alpha)}\right)^n \frac{e^{-\sum_{i=1}^{\ell}\sum_{j=1}^{n_i} \boldsymbol{x'_{ij}\beta}}}{(1+\sigma^2)^2\,(1+\alpha)^2} \\ &\quad\times \left[\sum_{k=1}^{m} e^{-(\alpha\, n_i\hat{z}_k\sigma + e^{-\hat{z}_k\sigma}\sum_{j=1}^{n_i} y_{ij}\, e^{-\boldsymbol{x'_{ij}\beta}})} \times (\Phi(t_k) - \Phi(t_{k-1}))\right].\end{aligned}$$

In the MH algorithm MCMC sequence, we keep the new sample only when it moves and we check the acceptance rate of the MH algorithm and test the convergence of the MCMC sequence.

(iv) Draw parameters $\nu_i|\boldsymbol{\beta}, \alpha, \sigma^2$ using the Metropolis–Hastings algorithm. The proposal density is a t-distribution with d degrees of freedom. We take mean and variance for the proposal from the samples of ν_i while drawing jointly $\boldsymbol{\beta}, \alpha$, and σ^2 in the above step. The target density is

$$\pi(\nu_i|\boldsymbol{\beta}, \sigma^2) \propto e^{-\left(\alpha\, n_i\nu_i + e^{-\nu_i}\sum_{j=1}^{n_i} y_{ij}\, e^{-\boldsymbol{x'_{ij}\beta}} + \frac{\nu_i^2}{2\sigma^2}\right)}, \quad i = 1, \cdots, \ell.$$

In the MH algorithm for random area effects, we keep the sample if its acceptance rate falls between 0.25 to 0.50. As mentioned before in the exponential model with random area effects, we do grid sampling for ν_i in the second attempt if its acceptance rate does not fall between 0.25 to 0.50.

2.4.2 Prediction

After drawing the set of parameters $\alpha, \boldsymbol{\beta}, \boldsymbol{\nu}$, and σ^2 from the hierarchical Bayesian gamma model with random area effects, we predict the responses as follows:

(i) Draw random area effects. For each PSU we have one random area effects parameter, $\nu_i, i = 1, \cdots, \ell$. We have sampled these parameters for all PSUs of NLSS-II. We will use these random area effects parameters to find the rate parameter.

(ii) Obtain the rate parameters directly as follows:

$$\lambda_{ij} = e^{-(\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\nu_i)}.$$

(iii) Draw predicted responses from the gamma distribution

$$\hat{y}_{ij} \sim \text{Gamma}(\alpha, \lambda_{ij}).$$

2.5 Generalized Gamma Model without Random Area Effects

The generalized gamma distribution has one more shape parameter than the gamma distribution, and it is a generalization of the gamma distribution. It has two shape parameters, α and γ; and one rate parameter, λ. We assume that the responses $y_i|\boldsymbol{\beta},\ i = 1, \cdots, n$ are random samples from the generalized gamma distribution with the rate $e^{-\boldsymbol{x}'_i\boldsymbol{\beta}}$. The likelihood function is

$$\begin{aligned} f(\boldsymbol{y}|\alpha, \boldsymbol{\beta}, \gamma) &= \prod_{i=1}^{n} \gamma \frac{e^{-y_i^{\gamma}\, e^{-\boldsymbol{x}'_i\boldsymbol{\beta}}}\, y_i^{\alpha\gamma-1}}{\Gamma(\alpha)}\, e^{-\alpha \boldsymbol{x}'_i\boldsymbol{\beta}} \\ &= \left(\frac{\gamma g^{\alpha\gamma-1}}{\Gamma(\alpha)}\right)^n e^{-\sum_{i=1}^{n}\left(\alpha\, \boldsymbol{x}'_i\boldsymbol{\beta} + y_i^{\gamma}\, e^{-\boldsymbol{x}'_i\boldsymbol{\beta}}\right)}, \end{aligned} \tag{2.33}$$

where $g = \left(\prod_{i=1}^{n} y_i\right)^{\frac{1}{n}}$ is the geometric mean of the response variable. We assume a non-informative prior for α and $\boldsymbol{\beta}$ and an informative prior for γ. The priors are independent.

The hierarchical Bayesian generalized gamma model without random area effects is

$$
\begin{aligned}
y_i|\alpha, \boldsymbol{\beta}, \gamma &\overset{\text{ind}}{\sim} \text{GGamma}(\alpha, e^{-\boldsymbol{x_i'\beta}}, \gamma), \quad \lambda_i = e^{-\boldsymbol{x_i'\beta}},\ i = 1, \cdots, n, \\
\pi(\boldsymbol{\beta}, \alpha) &\propto \frac{1}{(1+\alpha)^2} \\
\gamma &\sim \text{Gamma}(S, R), \quad \text{where shape S and rate R are specified.}
\end{aligned} \tag{2.34}
$$

Combining the likelihood in (2.33) and the priors in (2.34) via Bayes' theorem, we get the joint posterior density of $\alpha, \boldsymbol{\beta}$, and γ given sample data as

$$
\begin{aligned}
\pi(\alpha, \boldsymbol{\beta}, \gamma|\boldsymbol{y}) &\propto f(\boldsymbol{y}|\alpha, \boldsymbol{\beta}, \gamma)\, \pi(\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma) \\
&= \frac{e^{-R\gamma}\gamma^{S-1}}{(1+\alpha)^2} \left(\gamma \frac{g^{\alpha\gamma-1}}{\Gamma(\alpha)}\right)^n e^{-\sum_{i=1}^n \left(\alpha \boldsymbol{x_i'\beta} + y_i^\gamma\, e^{-\boldsymbol{x_i'\beta}}\right)}.
\end{aligned} \tag{2.35}
$$

Let the log-likelihood function be $G(\alpha, \boldsymbol{\beta}, \gamma|\boldsymbol{y}) = log(f(\boldsymbol{y}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}))$,

$$
G(\alpha, \boldsymbol{\beta}, \gamma|\boldsymbol{y}) = n\left[\log(\gamma) + (\alpha\,\gamma - 1)\log(g) - \log(\Gamma(\alpha))\right] - \sum_{i=1}^n \left(\alpha\, \boldsymbol{x_i'\beta} \;+\; y_i^\gamma\, e^{-\boldsymbol{x_i'\beta}}\right).
$$

For notational simplicity, let us write G for the log-likelihood function. Then its first- and second-order partial derivatives with respect to $\boldsymbol{\beta}$ are

$$
\frac{\partial G}{\partial \boldsymbol{\beta}} = -\sum_{i=1}^n \left(\alpha \;-\; y_i^\gamma\, e^{-\boldsymbol{x_i'\beta}}\right) \boldsymbol{x_i}, \quad \frac{\partial^2 G}{\partial \boldsymbol{\beta}^2} = -\sum_{i=1}^n \left(y_i^\gamma\, e^{-\boldsymbol{x_i'\beta}}\, \boldsymbol{x_i x_i'}\right).
$$

If we let $e^{-\boldsymbol{x_i'\beta}} \approx (1 - \boldsymbol{x_i'\beta})$, the first-order Taylor's series approximation at $\boldsymbol{\beta} = \boldsymbol{0}$. Then the approximated MLE of $\boldsymbol{\beta}|\alpha, \gamma$ is

$$
\boldsymbol{\beta}^*|\alpha, \gamma = \left[\sum_{i=1}^n y_i^\gamma (\boldsymbol{x_i x_i'})\right]^{-1} \left(\sum_{i=1}^n (y_i^\gamma - \alpha)\boldsymbol{x_i}\right). \tag{2.36}
$$

Let the gradient vector and the Hessian matrix evaluated at the approximate mode values $\boldsymbol{\beta}^*$ be $\nabla G_{\alpha\gamma}(\beta^*)$ and $H_{\alpha\gamma}(\boldsymbol{\beta}^*)$ respectively, then using the *multivariate normal approximation theorem*, we approximate the joint posterior density as

$$
\begin{aligned}
f(\alpha, \boldsymbol{\beta}, \gamma|\boldsymbol{y}) \approx\ & e^{\left[G_{\alpha\gamma}(\boldsymbol{\beta^*}) + \frac{1}{2}(\nabla G_{\alpha\gamma}(\boldsymbol{\beta^*}))'\,(-H_{\alpha\gamma}(\boldsymbol{\beta^*}))^{-1}\,\nabla G_{\alpha\gamma}(\boldsymbol{\beta^*})\right]} \\
& \times \left|(-H_{\alpha\gamma}(\boldsymbol{\beta^*}))^{-1}\right|^{\frac{1}{2}}\ N\left[\beta^* + (-H_{\alpha\gamma}(\boldsymbol{\beta^*}))^{-1}\,\nabla G_{\alpha\gamma}(\boldsymbol{\beta^*}),\ (-H_{\alpha\gamma}(\boldsymbol{\beta^*}))^{-1}\right] \\
& \times (2\pi)^{\frac{p+\ell}{2}}\, \frac{e^{-R\gamma}\,\gamma^{S-1}}{(1+\sigma^2)^2}.
\end{aligned}
$$

From the above joint distribution, we see that $\boldsymbol{\beta}$ has multivariate normal distribution as follows:

$$\boldsymbol{\beta}|\alpha, \gamma, \boldsymbol{y} \sim \mathrm{N}\left(\beta^* + (-H_{\alpha\gamma}(\boldsymbol{\beta}^*))^{-1}\,\nabla G_{\alpha\gamma}(\boldsymbol{\beta}^*),\ (-H_{\alpha\gamma}(\boldsymbol{\beta}^*))^{-1}\right). \tag{2.37}$$

Integrating out $\boldsymbol{\beta}$ we get the joint distribution of α and γ as

$$\begin{aligned}\pi(\alpha, \gamma|\boldsymbol{y}) &\approx e^{\left[G_{\alpha\gamma}(\boldsymbol{\beta}^*)+\frac{1}{2}(\nabla G_{\alpha\gamma}(\boldsymbol{\beta}^*))'\,(-H_{\alpha\gamma}(\boldsymbol{\beta}^*))^{-1}\,\nabla G_{\alpha\gamma}(\boldsymbol{\beta}^*)\right]} \times (2\pi)^{\frac{p+\ell}{2}}\ \left|(-H_{\alpha\gamma}(\boldsymbol{\beta}^*))^{-1}\right|^{\frac{1}{2}} \\ &\quad \times \frac{e^{-R\gamma}\,\gamma^{S-1}}{(1+\sigma^2)^2}.\end{aligned} \tag{2.38}$$

2.5.1 Sampling from Joint Posterior Density

(i) Draw α and γ jointly. We draw α and γ jointly using the two-dimensional grid method. Since $\alpha \in (0, \infty)$ and $\gamma \in (0, \infty)$, we transform them into $\eta = \frac{\alpha}{1+\alpha}$ and $\zeta = \frac{\gamma}{1+\gamma}$. We make 200×200 grids for η and ζ and the calculate joint probability density of transformed density(2.38). We draw a set of 200 samples jointly with replacement.

(ii) Using the information from α and γ, we could draw $\boldsymbol{\beta}|\alpha, \gamma, \boldsymbol{y}$. The Metropolis–Hastings algorithm is used to draw $\boldsymbol{\beta}, \alpha, \gamma|\boldsymbol{y}$ jointly. The proposal densities are t-distributions. We take the log-transformation for the joint proposal densities α and γ. Then consider $\log(\alpha, \gamma)|\boldsymbol{y}$ as the bivariate normal, $[\log(\alpha), \log(\gamma)]|\boldsymbol{y} \sim \mathrm{N}(\boldsymbol{\mu}_{ln}, \Sigma_{ln})$, where $\boldsymbol{\mu}_{ln}$ and Σ_{ln} are calculated from the α and γ samples drawn in the above step. The proposal density function for $\boldsymbol{\beta}|\alpha, \gamma$ is the multivariate t-distribution with d degrees of freedom, with the corresponding mean and covariance matrix as in (2.37). The target density is the posterior density (2.35).

2.5.2 Prediction

After drawing samples from the generalized gamma model, we can predict the responses as follows:

(i) Find the rate parameter $\lambda_i = e^{-\boldsymbol{x}_i'\boldsymbol{\beta}}$.

(ii) In the generalized gamma distribution, λy^{γ} has the gamma distribution with rate unity and shape α. Let us say we draw a random sample G_1 from the gamma distribution as follows:

$$\mathrm{G}_1 = (\lambda_i y)^{\gamma} \sim \text{Gamma}(\alpha, 1).$$

We can predict the response variable as follows:

$$\hat{y}_i = \frac{\mathrm{G}_1^{\frac{1}{\gamma}}}{\lambda_i}.$$

2.6 Generalized Gamma Model with Random Area Effects

The generalized gamma distribution is the generalization of the gamma distribution. It has one more shape parameter γ than the gamma distribution. If $\gamma = 1$, then the generalized gamma distribution becomes the gamma distribution. We assume the responses $y_{ij}|\alpha, \boldsymbol{\beta}, \gamma$ are independent random samples from the generalized gamma distribution with rate $e^{-(\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\nu_i)}$. We assume that ν_i follow the normal distribution with mean zero and variance σ^2. The likelihood function is

$$\begin{aligned} f(\boldsymbol{y}|\alpha, \boldsymbol{\beta}, \gamma, \boldsymbol{\nu}) &= \prod_{i=1}^{\ell}\prod_{j=1}^{n_i} \gamma \frac{e^{-y_{ij}^{\gamma}\, e^{-(\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\nu_i)}}\, y_{ij}^{\alpha\,\gamma-1}}{\Gamma(\alpha)}\, e^{-\alpha(\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\nu_i)} \\ &= \left(\frac{\gamma\, g^{\alpha\gamma-1}}{\Gamma(\alpha)}\right)^{n} e^{-\alpha\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\boldsymbol{x}'_{ij}\boldsymbol{\beta}}\, e^{-\sum_{i=1}^{\ell}\left(\alpha n_i\nu_i + e^{-\nu_i}\sum_{j=1}^{n_i} y_{ij}^{\gamma} e^{-\boldsymbol{x}'_{ij}\boldsymbol{\beta}}\right)}, \end{aligned} \tag{2.39}$$

where $g = \left(\prod_{i=1}^{\ell}\prod_{j=1}^{n_i} y_{ij}\right)^{\frac{1}{n}}$, the geometric mean of the response variable. We assume that $\alpha, \boldsymbol{\beta}$, and σ^2 have non-informative priors and γ has an informative prior. The priors are independent. The hierarchical Bayesian generalized gamma model with random area effects is

$$
\begin{aligned}
y_{ij}|\alpha,\boldsymbol{\beta},\gamma,\nu_i &\overset{\text{ind}}{\sim} \text{GGamma}\left(\alpha, e^{-(\boldsymbol{x_{ij}'}\boldsymbol{\beta}+\boldsymbol{\nu_i})},\gamma\right), \quad \lambda_{ij}=e^{-(\boldsymbol{x_{ij}'}\boldsymbol{\beta}+\boldsymbol{\nu_i})}, \qquad (2.40)\\
\nu_i &\overset{\text{iid}}{\sim} N(0,\sigma^2), \quad i=1,\cdots,\ell,\ j=1,\cdots,n_i,\\
\pi(\boldsymbol{\beta},\alpha,\sigma^2) &\propto \frac{1}{(1+\alpha^2)(1+\sigma^2)},\\
\gamma &\sim \text{Gamma}(S,R), \quad \text{where shape S and rate R are specified.} \qquad (2.41)
\end{aligned}
$$

Combining the likelihood in (2.39) and the priors in (4.3) via Bayes' theorem, we get the joint posterior density of $\alpha,\boldsymbol{\beta},\gamma$, and σ^2, given sample data as

$$
\begin{aligned}
&\pi(\alpha,\boldsymbol{\beta},\gamma,\boldsymbol{\nu},\sigma^2|\boldsymbol{y}) \propto f(\boldsymbol{y}|\boldsymbol{\beta},\alpha)\,\pi(\boldsymbol{\beta},\alpha,\gamma,\sigma^2)\,\pi(\boldsymbol{\nu})\\
&=\left(\gamma\,\tfrac{g^{\alpha\gamma-1}}{\Gamma(\alpha)}\right)^n e^{-\alpha\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\boldsymbol{x_{ij}'}\boldsymbol{\beta}}\, e^{-\sum_{i=1}^{\ell}\left(\alpha n_i\nu_i+e^{-\nu_i}\sum_{j=1}^{n_i}y_{ij}^{\gamma}e^{-\boldsymbol{x_{ij}'}\boldsymbol{\beta}}\right)}\times\prod_{i=1}^{\ell}\left[\left(\tfrac{1}{\sigma^2}\right)^{\frac{1}{2}}e^{-\frac{\nu_i^2}{2\sigma^2}}\right]\tfrac{e^{-R\gamma}\gamma^{S-1}}{(1+\sigma^2)^2\,(1+\alpha)^2}\\
&=\left(\gamma\,\tfrac{g^{\alpha\gamma-1}}{\Gamma(\alpha)}\right)^n e^{-\alpha\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\boldsymbol{x_{ij}'}\boldsymbol{\beta}}\, e^{-\sum_{i=1}^{\ell}\left(\alpha n_i\nu_i+e^{-\nu_i}\sum_{j=1}^{n_i}y_{ij}^{\gamma}e^{-\boldsymbol{x_{ij}'}\boldsymbol{\beta}}+\frac{\nu_i^2}{2\sigma^2}\right)}\times\tfrac{e^{-R\gamma}\gamma^{S-1}}{(1+\sigma^2)^2\,(1+\alpha)^2}\left(\tfrac{1}{\sigma^2}\right)^{\frac{\ell}{2}}. \qquad (2.42)
\end{aligned}
$$

Let the log-likelihood function be $G(\alpha,\boldsymbol{\tau},\gamma|\boldsymbol{y})=log(f(\boldsymbol{y|\alpha,\tau,\gamma}))$, where $\boldsymbol{\tau}=(\boldsymbol{\beta'},\,\boldsymbol{\nu'})'$,

$$
\begin{aligned}
&G(\alpha,\boldsymbol{\tau},\gamma|\boldsymbol{y})\\
&=n\left[\log(\gamma)+(\alpha\,\gamma-1)\log(g)-\log(\Gamma(\alpha))\right]-\alpha\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\boldsymbol{x_{ij}'}\boldsymbol{\beta}\sum_{i=1}^{\ell}\left(\alpha n_i\nu_i\,+\,e^{-\nu_i}\sum_{j=1}^{n_i}y_{ij}^{\gamma}e^{-\boldsymbol{x_{ij}'}\boldsymbol{\beta}}\right).
\end{aligned}
$$

For notational simplicity, let us write G for the log-likelihood function. Then its first- and second-order partial derivatives with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\nu}$ are

$$
\begin{aligned}
\frac{\partial G}{\partial\boldsymbol{\beta}} &= -\,\alpha\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\boldsymbol{x_{ij}}+\sum_{i=1}^{\ell}e^{-\nu_i}\sum_{j=1}^{n_i}y_{ij}^{\gamma}\,e^{-\boldsymbol{x_{ij}'}\boldsymbol{\beta}}\,\boldsymbol{x_{ij}},\\
\frac{\partial G}{\partial\nu_i} &= -\left(\alpha\,n_i-e^{-\nu_i}\sum_{j=1}^{n_i}y_{ij}^{\gamma}\,e^{-\boldsymbol{x_{ij}'}\boldsymbol{\beta}}\right),\\
\frac{\partial^2 G}{\partial\boldsymbol{\beta}^2} &= -\sum_{i=1}^{\ell}e^{-\nu_i}\sum_{j=1}^{n_i}y_{ij}^{\gamma}\,e^{-\boldsymbol{x_{ij}'}\boldsymbol{\beta}}\,\boldsymbol{x_{ij}}\boldsymbol{x_{ij}'},\\
\frac{\partial^2 G}{\partial\nu_i^2} &= -\,e^{-\nu_i}\sum_{j=1}^{n_i}y_{ij}^{\gamma}\,e^{-\boldsymbol{x_{ij}'}\boldsymbol{\beta}},\\
\frac{\partial^2 G}{\partial\boldsymbol{\beta}\partial\nu_i} &= -\,e^{-\nu_i}\sum_{j=1}^{n_i}y_{ij}^{\gamma}\,e^{-\boldsymbol{x_{ij}'}\boldsymbol{\beta}}\,\boldsymbol{x_{ij}}.
\end{aligned}
$$

Let $e^{-\boldsymbol{x_i'}\boldsymbol{\beta}}\approx(1-\boldsymbol{x_i'}\boldsymbol{\beta})$, the first-order Taylor's series approximation at $\boldsymbol{\beta}=\boldsymbol{0}$, then we have the approximated MLE of $\boldsymbol{\beta}|\alpha,\gamma$ as

$$\boldsymbol{\beta}^*|\alpha,\gamma = \left[\sum_{i=1}^{\ell} e^{-\nu_i} \sum_{j=1}^{n_i} y_{ij}^{\gamma}(\boldsymbol{x_i x_i'})\right]^{-1} \left(-\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}(e^{-\nu_i}y_{ij}^{\gamma}-\alpha)\boldsymbol{x_{ij}'}\right). \tag{2.43}$$

The approximated MLE of ν_i is given by

$$\boldsymbol{\nu_i^*} = -\log\left[\frac{\alpha\, n_i}{\sum_{j=1}^{n_i} y_{ij}^{\gamma}\, e^{-\boldsymbol{x_{ij}'\beta}}}\right]. \tag{2.44}$$

Let the gradient vectors be $\nabla G_{\alpha\gamma}(\tau^*) = (\boldsymbol{g_\nu'},\, \boldsymbol{g_\beta'})'$, where $\boldsymbol{g_\nu} = \left(\frac{\partial G}{\partial \nu_1} \;\cdots\; \frac{\partial G}{\partial \nu_\ell}\right)'|_{\boldsymbol{\nu}=\boldsymbol{\nu}^*,\,\boldsymbol{\beta}=\boldsymbol{\beta}^*}$, and $\boldsymbol{g_\beta} = \left(\frac{\partial G}{\partial \beta_0} \;\cdots\; \frac{\partial G}{\partial \beta_p}\right)'|_{\boldsymbol{\nu}=\boldsymbol{\nu}^*,\,\boldsymbol{\beta}=\boldsymbol{\beta}^*}$, and the Hessian matrix be $H_{\alpha\gamma}(\tau^*)$ evaluated at the approximate mode values $\boldsymbol{\beta}^*$ and $\boldsymbol{\nu}^*$. Then applying the second-order Taylor's series approximation, we can write the approximated likelihood function as

$$\begin{aligned} f(\boldsymbol{y}|\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\nu}) &\approx e^{\left[G_{\alpha\gamma}(\boldsymbol{\tau^*})+\frac{1}{2}(\nabla G_{\alpha\gamma}(\tau^*))'\,(-H_{\alpha\gamma}(\boldsymbol{\tau^*}))^{-1}\,\nabla G_{\alpha\gamma}(\tau^*)\right]} \\ &\quad (2\pi)^{\frac{p+\ell}{2}}\, \left|(-H_{\alpha\gamma}(\tau^*))^{-1}\right|^{\frac{1}{2}}\; N\left[\tau^* + (-H_{\alpha\gamma}(\boldsymbol{\tau^*}))^{-1}\,\nabla G_{\alpha\gamma}(\tau^*),\; (-H_{\alpha\gamma}(\boldsymbol{\tau^*}))^{-1}\right], \end{aligned}$$

where N denotes the multivariate normal distribution for the parameter set $\boldsymbol{\tau} = (\boldsymbol{\beta'},\, \boldsymbol{\nu'})'$. Following the *multivariate normal approximation theorem* in Chapter 1 we can write

$$\begin{pmatrix}\boldsymbol{\nu}\\ \boldsymbol{\beta}\end{pmatrix} \sim N\left\{\begin{pmatrix}\boldsymbol{\mu_\nu^*}\\ \boldsymbol{\mu_\beta^*}\end{pmatrix}, \begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{12}' & \Sigma_{22}\end{pmatrix}\right\},$$

where the Hessian matrix is $H = -\begin{pmatrix}A_{11} & A_{12}\\ A_{12}' & A_{22}\end{pmatrix}$. Using the same notation as in Chapter 1, equations 1.3 and 1.4 for vectors and matrices and then applying the *multivariate normal approximation theorem*, we can write the approximated joint posterior density as

$$\begin{aligned} & f(\boldsymbol{\beta},\boldsymbol{\nu},\alpha,\gamma,\sigma^2|\boldsymbol{y}) \\ & \propto C_{\alpha\gamma}(\tau^*) \times N\left(\boldsymbol{\mu}_\beta^*,\; \Sigma_{22}\right) \times N\left(\boldsymbol{\mu}_\nu^* + \Sigma_{12}\Sigma_{22}^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu}_\beta^*),\; \Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}'\right) \times N\left(\boldsymbol{0},\, \sigma^2 I_\ell\right) \\ & \quad \times \frac{e^{-R\gamma}\gamma^{S-1}}{(1+\sigma^2)^2 1+\alpha^2} \\ & = C_{\alpha\gamma}(\tau^*) \times \frac{e^{-R\gamma}\gamma^{S-1}}{(1+\sigma^2)^2\,(1+\alpha)^2} \times \frac{|A_{11}|^{\frac{1}{2}}}{|\Sigma_{22}|^{\frac{1}{2}}\,|\sigma^2 I_\ell|^{\frac{1}{2}}} \times e^{-\frac{1}{2}\left[(\boldsymbol{\beta}-\boldsymbol{\mu}_\beta^*)'\,\Sigma_{22}^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu}_\beta^*)\right]} \\ & \quad \times e^{-\frac{1}{2}\left[(\boldsymbol{\mu}_\nu^*-A_{11}^{-1}A_{12}(\boldsymbol{\beta}-\boldsymbol{\mu}_\beta^*))'A_{11}\left(\left(A_{11}+(\sigma^2 I_\ell\right)^{-1}\right)^{-1}\left(\sigma^2 I_\ell\right)^{-1}(\boldsymbol{\mu}_\nu^*-A_{11}^{-1}A_{12}(\boldsymbol{\beta}-\boldsymbol{\mu}_\beta^*))\right]} \\ & \quad \times e^{-\frac{1}{2}\left[\left[\boldsymbol{\nu}-\left(A_{11}+(\sigma^2 I_\ell)^{-1}\right)^{-1}(A_{11}\boldsymbol{\mu}_\nu^*-A_{12}(\boldsymbol{\beta}-\boldsymbol{\mu}_\beta^*))\right]'\left(A_{11}+(\sigma^2 I_\ell)^{-1}\right)\left[\boldsymbol{\nu}-\left(A_{11}+(\sigma^2 I_\ell)^{-1}\right)^{-1}(A_{11}\boldsymbol{\mu}_\nu^*-A_{12}(\boldsymbol{\beta}-\boldsymbol{\mu}_\beta^*))\right]\right]}, \end{aligned} \tag{2.45}$$

where

$$C_{\alpha\gamma}(\tau^*) = e^{\left[G_{\alpha\gamma}(\boldsymbol{\tau}^*)+\frac{1}{2}(\nabla G_{\alpha\gamma}(\tau^*))'\,(-H_{\alpha\gamma}(\boldsymbol{\tau}^*))^{-1}\,\nabla G_{\alpha\gamma}(\tau^*)\right]}\left|(-H_{\alpha\gamma}(\tau^*))^{-1}\right|^{\frac{1}{2}}.$$

From the above joint posterior density, we notice that $\boldsymbol{\nu}$ has the multivariate normal distribution

$$\boldsymbol{\nu}|\boldsymbol{\beta},\sigma^2 \sim N\left[\left(A_{11}+(\sigma^2 I_\ell)^{-1}\right)^{-1}(A_{11}\boldsymbol{\mu}_\nu^* - A_{12}(\boldsymbol{\beta}-\boldsymbol{\mu}_\beta^*)),\ \left(A_{11}+(\sigma^2 I_\ell)^{-1}\right)^{-1}\right]. \quad (2.46)$$

As before, we integrate ν_i from the joint posterior density. Integrating out $\boldsymbol{\nu}$ we have a joint distribution of $\boldsymbol{\beta},\alpha,\gamma,\sigma^2|\boldsymbol{y}$ as

$$\begin{aligned}
&f(\boldsymbol{\beta},\alpha,\gamma,\sigma^2|\boldsymbol{y})\\
&\propto C_{\alpha\gamma}(\tau^*)\times\frac{e^{-R\gamma}\gamma^{S-1}}{(1+\sigma^2)^2\,(1+\alpha)^2}\times\frac{|A_{11}|^{\frac{1}{2}}}{|\Sigma_{22}|^{\frac{1}{2}}\ |\sigma^2 I_\ell|^{\frac{1}{2}}}\times e^{-\frac{1}{2}\left[(\boldsymbol{\beta}-\boldsymbol{\mu}_\beta^*)'\,\Sigma_{22}^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu}_\beta^*)\right]}\times\left|A_{11}+(\sigma^2 I_\ell)^{-1}\right|^{-\frac{1}{2}}\\
&\quad\times e^{-\frac{1}{2}\left[(\boldsymbol{\beta}-\widetilde{\boldsymbol{\mu}}_\beta)'\widetilde{\Sigma}(\boldsymbol{\beta}-\widetilde{\boldsymbol{\mu}}_\beta)-\widetilde{\boldsymbol{\mu}}'_\beta\widetilde{\Sigma}\widetilde{\boldsymbol{\mu}}_\beta+(\boldsymbol{\mu}_\nu^*+A_{11}^{-1}A_{12}\boldsymbol{\mu}_\beta^*)'S(\boldsymbol{\mu}_\nu^*+A_{11}^{-1}A_{12}\boldsymbol{\mu}_\beta^*)\right]},
\end{aligned}$$

where

$$\begin{aligned}
S &= A_{11}\left(A_{11}+(\sigma^2 I_\ell)^{-1}\right)^{-1}(\sigma^2 I_\ell)^{-1},\\
\widetilde{\boldsymbol{\mu}}_\beta &= (A'_{12}A_{11}^{-1}SA_{11}^{-1}A_{12})^{-1}A'_{12}A_{11}^{-1}S\boldsymbol{\mu}_\nu^*+\boldsymbol{\mu}_\beta^*,\\
\widetilde{\Sigma}_\beta &= A'_{12}A_{11}^{-1}SA_{11}^{-1}A_{12}.
\end{aligned}$$

From this joint posterior density, we notice that $\boldsymbol{\beta}$ has the multivariate normal distribution

$$\boldsymbol{\beta}|\alpha,\gamma,\sigma^2,\boldsymbol{y}\sim N\left[\left(\Sigma_{22}^{-1}+\widetilde{\Sigma}_\beta\right)^{-1}\left(\Sigma_{22}^{-1}\boldsymbol{\mu}_\beta^*+\widetilde{\Sigma}_\beta\widetilde{\boldsymbol{\mu}}_\beta\right),\ \left(\Sigma_{22}^{-1}+\widetilde{\Sigma}_\beta\right)^{-1}\right]. \quad (2.47)$$

Integrating out $\boldsymbol{\beta}$ we get the joint distribution of $(\alpha,\gamma,\sigma^2|\boldsymbol{y})$ as follows:

$$\begin{aligned}
&\pi(\alpha,\gamma,\sigma^2|\boldsymbol{y})\\
&\qquad\propto C_{\alpha\gamma}(\tau^*)\times\frac{e^{-R\gamma}\gamma^{S-1}}{(1+\sigma^2)^2\,(1+\alpha)^2}\,\frac{|A_{11}+(\sigma^2 I_\ell)^{-1}|^{-\frac{1}{2}}\left|\Sigma_{22}^{-1}+\widetilde{\Sigma}_\beta\right|^{-\frac{1}{2}}}{|\sigma^2 I_\ell|^{\frac{1}{2}}} \qquad (2.48)\\
&\qquad\quad\times e^{-\frac{1}{2}\left[(\boldsymbol{\mu}_\beta^*-\widetilde{\boldsymbol{\mu}}_\beta)'\Sigma_{22}^{-1}\left(\Sigma_{22}^{-1}+\widetilde{\Sigma}_\beta\right)^{-1}\widetilde{\Sigma}_\beta(\boldsymbol{\mu}_\beta^*-\widetilde{\boldsymbol{\mu}}_\beta)\right]}\,e^{-\frac{1}{2}\left[-\widetilde{\boldsymbol{\mu}}'_\beta\widetilde{\Sigma}_\beta\widetilde{\boldsymbol{\mu}}_\beta+(\boldsymbol{\mu}_\nu^*+A_{11}^{-1}A_{12}\boldsymbol{\mu}_\beta^*)'S(\boldsymbol{\mu}_\nu^*+A_{11}^{-1}A_{12}\boldsymbol{\mu}_\beta^*)\right]}.
\end{aligned}$$

2.6.1 Sampling from Joint Posterior Density

The joint posterior density of α,γ, and σ^2 is not in a simple form. We borrow parameters α and γ from the previous hierarchical Bayesian model without random area effects and used them as an approximation. We can draw $\boldsymbol{\beta}$ and $\boldsymbol{\nu}$ from the approximated multivariate normal distribution. However, σ^2 is not in simple form. We have used the grid method and MH algorithm methods to draw parameters.

(i) Borrow α and γ from the previous model. We borrow α and γ parameters sampled in the previous hierarchical Bayesian generalized gamma model without random area effects and use them in this model as an approximation. We choose a set of 100 quantiles from the total of 1,000 samples of α and γ from the previous model, the generalized gamma model without random area effect.

(ii) We draw $\sigma^2|\alpha, \gamma, \boldsymbol{y}$ using the grid sampling method with the density function given by

$$\begin{aligned}\pi(\sigma^2|\alpha,\gamma,\boldsymbol{y}) \propto \frac{1}{(1+\sigma^2)^2} & \frac{|A_{11}+(\sigma^2 I_\ell)^{-1}|^{-\frac{1}{2}} \left|\Sigma_{22}^{-1}+\widetilde{\Sigma}_\beta\right|^{-\frac{1}{2}}}{|\sigma^2 I_\ell|^{\frac{1}{2}}} \\ & \times e^{-\frac{1}{2}\left[(\boldsymbol{\mu}_\beta^* - \widetilde{\boldsymbol{\mu}}_\beta)'\Sigma_{22}^{-1}\left(\Sigma_{22}^{-1}+\widetilde{\Sigma}_\beta\right)^{-1}\widetilde{\Sigma}_\beta(\boldsymbol{\mu}_\beta^* - \widetilde{\boldsymbol{\mu}}_\beta)\right]} \\ & \times e^{-\frac{1}{2}\left[-\widetilde{\boldsymbol{\mu}}_\beta'\widetilde{\Sigma}_\beta\widetilde{\boldsymbol{\mu}}_\beta + (\boldsymbol{\mu}_\nu^* + A_{11}^{-1}A_{12}\boldsymbol{\mu}_\beta^*)'S(\boldsymbol{\mu}_\nu^* + A_{11}^{-1}A_{12}\boldsymbol{\mu}_\beta^*)\right]}. \end{aligned} \tag{2.49}$$

Since $\sigma^2 \in (0, \infty)$, we transform σ^2 into η, which has range $(0, 1)$, using $\eta = \frac{\sigma^2}{1+\sigma^2}$. We take 100 grids of η and compute the transformed probability $\pi(\eta|\alpha, \gamma, \boldsymbol{y})$ using (2.49). We draw samples from this grid probability distribution of $\eta|\alpha, \gamma, \boldsymbol{y}$ then transform it into σ^2.

(iii) Using the information in parameters α, γ and σ^2, we can draw $\boldsymbol{\beta}$. The Metropolis–Hastings algorithm is used for sampling jointly $\alpha, \boldsymbol{\beta}, \gamma$, and σ^2. The proposal densities are t-distributions. We take the log-transformation for the joint proposal density of the α, γ, and σ^2. Then we consider $[\log(\alpha), \log(\gamma), \log(\sigma^2)] \sim t_d(\boldsymbol{\mu}_{ln}, \Sigma_{ln})$, where $\boldsymbol{\mu}_{ln}$ and Σ_{ln} are estimated from the previous step's samples of α, γ and σ^2. The proposal distribution for $\boldsymbol{\beta}|\alpha, \gamma, \sigma^2, \boldsymbol{y}$ is the multivariate t-distribution with d degrees of freedom, with corresponding mean and covariance matrix as in equation (2.47).

The target density is

$$\pi(\boldsymbol{\beta},\alpha,\gamma,\sigma^2|\boldsymbol{y}) \propto \left(\frac{\gamma\, g^{\alpha\gamma-1}}{\Gamma(\alpha)}\right)^n \frac{e^{-R\gamma}\gamma^{S-1}}{(1+\sigma^2)^2\,(1+\alpha)^2}\, e^{-\alpha\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\boldsymbol{x'_{ij}\beta}} \times \prod_{i=1}^{\ell}\left[\int_{\nu_i} e^{-(\alpha\, n_i\nu_i + e^{-\nu_i}\sum_{j=1}^{n_i} y_{ij}^{\gamma}\, e^{-\boldsymbol{x'_{ij}\beta}} + \frac{\nu_i^2}{2\sigma^2})} \times \left(\frac{1}{\sigma^2}\right)^{\frac{1}{2}} d\nu_i\right].$$

This integration is not in simple form. We perform numerical integration. We divide the integration domain into m equal intervals $[t_k,\, t_{k-1}]$.

$$\pi(\boldsymbol{\beta},\alpha,\gamma,\sigma^2|\boldsymbol{y}) \propto \left(\frac{\gamma\, g^{\alpha\gamma-1}}{\Gamma(\alpha)}\right)^n \frac{e^{-R\gamma}\gamma^{S-1}}{(1+\sigma^2)^2\,(1+\alpha)^2}\, e^{-\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\boldsymbol{x'_{ij}\beta}} \times \prod_{i=1}^{\ell}\left[\sum_{k=1}^{m}\int_{t_{k-1}}^{t_k} e^{-(\alpha\, n_i\nu_i + e^{-\nu_i}\sum_{j=1}^{n_i} y_{ij}^{\gamma}\, e^{-\boldsymbol{x'_{ij}\beta}})} \times \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\nu_i^2}{2\sigma^2}} d\nu_i\right].$$

Using the assumption of ν_i having a univariate normal distribution, we transform ν_i to the standard normal distribution, $z_i = \frac{\nu_i}{\sigma}$. For numerical integration we approximate by taking the middle point of each interval $\hat{z}_k = \frac{t_{k-1}+t_k}{2}$,

$$\begin{aligned}\pi(\boldsymbol{\beta},\alpha,\gamma,\sigma^2|\boldsymbol{y}) \propto\ & \left(\frac{\gamma\, g^{\alpha\gamma-1}}{\Gamma(\alpha)}\right)^n \frac{e^{-R\gamma}\gamma^{S-1}}{(1+\sigma^2)^2\,(1+\alpha)^2}\, e^{-\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\boldsymbol{x'_{ij}\beta}} \\ & \times \prod_{i=1}^{\ell}\left[\sum_{k=1}^{m} e^{-(\alpha\, n_i\hat{z}_k\sigma + e^{-\hat{z}_k\sigma}\sum_{j=1}^{n_i} y_{ij}^{\gamma}\, e^{-\boldsymbol{x'_{ij}\beta}})} \times \int_{t_{k-1}}^{t_k}\frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz\right] \\ =\ & \left(\frac{\gamma\, g^{\alpha\gamma-1}}{\Gamma(\alpha)}\right)^n \frac{e^{-R\gamma}\gamma^{S-1}}{(1+\sigma^2)^2\,(1+\alpha)^2}\frac{1}{-}\, e^{-\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\boldsymbol{x'_{ij}\beta}} \\ & \times \prod_{i=1}^{\ell}\left[\sum_{k=1}^{m} e^{-(\alpha\, n_i\hat{z}_k\sigma + e^{-\hat{z}_k\sigma}\sum_{j=1}^{n_i} y_{ij}^{\gamma}\, e^{-\boldsymbol{x'_{ij}\beta}})} \times (\Phi(t_k) - \Phi(t_{k-1}))\right].\end{aligned}$$

(iv) Draw ν_i. Parameters $\nu_i|\boldsymbol{\beta},\alpha,\gamma,\sigma^2$ are drawn using the Metropolis–Hastings algorithm. The proposal density is the t-distribution with d degrees of freedom. We take the mean and variance for the proposal from the previous step's samples of ν_i while drawing $\boldsymbol{\beta},\alpha,\gamma$, and σ^2 jointly.

$$\pi(\nu_i|\boldsymbol{\beta},\alpha,\gamma,\sigma^2) \propto e^{-\left(\alpha\, n_i\nu_i + e^{-\nu_i}\sum_{j=1}^{n_i} y_{ij}^{\gamma}\, e^{-\boldsymbol{x'_{ij}\beta}} + \frac{\nu_i^2}{2\sigma^2}\right)}, \quad i = 1,\cdots,\ell.$$

The acceptance rate of the random area effect can go much larger than 0.50 and much smaller than 0.25 so we keep the sample ν_i from the Metropolis–Hastings sampler if its acceptance rate is between 0.25 and 0.50; if not then we discard this sample. In the next attempt we perform grid sampling to draw the same random area effect.

2.6.2 Prediction

After drawing all sets of parameters from the generalized gamma distribution model as mentioned above, we predict the responses as follows:

(i) Draw the random area effect. For each PSU we have one random area effects parameter, $\nu_i, i = 1, \cdots, \ell$. We have sampled these parameters for all PSUs of NLSS-II. Using these sampled random area effects parameters, we can obtain the rate parameters directly by

$$\lambda_{ij} = e^{-(\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\nu_i)}.$$

(ii) Draw a predicted response from the generalized gamma distribution. Consider the transformation $t = y^{\gamma}$. This gives

$$\begin{aligned} G &= (y_{ij})^{\gamma} \sim \text{Gamma}(\alpha, \lambda_{ij}) \quad \text{and} \\ \hat{y}_{ij} &= G^{\frac{1}{\gamma}}. \end{aligned}$$

2.7 Model Validation and Model comparison

One way to evaluate the adequacy of a model is by Bayesian cross-validation. We consider the cross-validation approach by Gelfand, Dey, and Chang (1992). The cross-validation approach involves the prediction of subset y_i of the response data $\boldsymbol{y}$, when only the component $\boldsymbol{y}_{(i)}$ is used. Let $\boldsymbol{y}$ be the data vector of $N \times 1$ and $\boldsymbol{y}_{(i)}$ and denote $(N-1) \times 1$ data vector with the i^{th} observation deleted. If we fit a model with $\boldsymbol{y}_{(i)}$ and if the model fits well, then it should predict y_i very well. The prediction of y_i is $E(Y_{ij}|\boldsymbol{y}_{(ij)})$ with variance $Var(y_i|\boldsymbol{y}_{(i)})$. The cross-validation standardized deleted residuals are

$$r_{ij} = \frac{Y_{ij} - E(Y_{ij}|\boldsymbol{y}_{(ij)})}{\sqrt{Var(y_i|\boldsymbol{y}_{(i)})}}, \qquad i = 1, \cdots, \ell, j = 1, \cdots, n_i. \tag{2.50}$$

We can approximate the posterior expectation and variance in (2.50) by weighted average.

$$\begin{aligned} E(Y_{ij}|\boldsymbol{y}_{(ij)}) &= \int_{\boldsymbol{\Omega}} E(Y_{ij}|\boldsymbol{y}_{(ij)}, \boldsymbol{\Omega}) \ \pi(\boldsymbol{\Omega}|\boldsymbol{y}_{ij}) \, d\boldsymbol{\Omega} \\ &\approx \sum_{h=1}^{M} \hat{y}_{ij}^{(h)} \ w_{ij}^{(h)}, \end{aligned}$$

where Ω the parameters set and weights are given by

$$w_{ij}^{(h)} = \left[f(Y_{ij}|\boldsymbol{\Omega}^{(h)}) \sum_{h=1}^{M} \left(f(Y_{ij}|\boldsymbol{\Omega}^{(h)}) \right)^{-1} \right]^{-1}.$$

The cross-validation approach needs to find $p(y_i|y_{(i)})$, called the cross-validation predictive distribution or conditional predictive distribution (CPO). We consider the standardized cross-validation residuals and conditional predictive ordinates defined by Box (1980) and the studies under normal distribution by Pettit (1990).

$$\begin{aligned} CPO_i &= \int_{\boldsymbol{\Omega}} f(y_i|y_{(i)}, \boldsymbol{\Omega}) \, f(\boldsymbol{\Omega}|y_{(i)}) \, d\boldsymbol{\Omega} = E_{\boldsymbol{\Omega}|y_{(i)}}[p(y_i|\boldsymbol{\Omega})] \\ &= \frac{\int_{\boldsymbol{\Omega}} f(y_i|y_{(i)}, \boldsymbol{\Omega}) \, \frac{f(\boldsymbol{\Omega}|y_{(i)})}{f(\boldsymbol{\Omega}|\boldsymbol{y})} \, f(\boldsymbol{\Omega}|\boldsymbol{y}) \, d\boldsymbol{\Omega}}{\int_{\boldsymbol{\Omega}} \frac{f(\boldsymbol{\Omega}|y_{(i)})}{f(\boldsymbol{\Omega}|\boldsymbol{y})} \, f(\boldsymbol{\Omega}|\boldsymbol{y}) \, d\boldsymbol{\Omega}} \\ &\approx \frac{\sum_{k=1}^{M} f(y_i|y_{(i)}, \boldsymbol{\Omega}) \, \frac{f(\boldsymbol{\Omega}|y_{(i)})}{f(\boldsymbol{\Omega}|\boldsymbol{y})}}{\sum_{k=1}^{M} \frac{f(\boldsymbol{\Omega}|y_{(i)})}{f(\boldsymbol{\Omega}|\boldsymbol{y})}} \\ &= \sum_{k=1}^{M} f(y_i|y_{(i)}, \boldsymbol{\Omega}) \ w_i^{(k)}, \end{aligned}$$

where $w_i^{(k)} = \dfrac{\frac{f(\boldsymbol{\Omega}|y_{(i)})}{f(\boldsymbol{\Omega}|\boldsymbol{y})}}{\sum_{k=1}^{M} \frac{f(\boldsymbol{\Omega}|y_{(i)})}{f(\boldsymbol{\Omega}|\boldsymbol{y})}}$.

By Bayes' theorem,

$$\frac{f(\boldsymbol{\Omega}|y_{(i)})}{f(\boldsymbol{\Omega}|\boldsymbol{y})} = \frac{\frac{f(y_{(i)}|\boldsymbol{\Omega})}{f(\boldsymbol{y}|\boldsymbol{\Omega})}}{\frac{f(y_{(i)})}{f(\boldsymbol{y})}} \qquad \text{then} \qquad w_i^{(k)} = \frac{\frac{f(y_{(i)}|\boldsymbol{\Omega}^{(k)})}{f(\boldsymbol{y}|\boldsymbol{\Omega}^{(k)})}}{\sum_{k=1}^{M} \frac{f(y_{(i)}|\boldsymbol{\Omega}^{(k)})}{f(\boldsymbol{y}|\boldsymbol{\Omega}^{(k)})}} = \frac{\left[f(y_{(i)}|\boldsymbol{\Omega}^{(k)}) \right]^{-1}}{\sum_{k=1}^{M} \left[f(y_{(i)}|\Omega^{(k)}) \right]^{-1}}.$$

Therefore, we can write the conditional predictive density as

$$CPO_i \approx \sum_{k=1}^{M} f(y_i|y_{(i)}, \boldsymbol{\Omega}) \left[\frac{\left[f(y_{(i)}|\boldsymbol{\Omega}^{(k)}) \right]^{-1}}{\sum_{k=1}^{M} \left[f(y_{(i)}|\boldsymbol{\Omega}^{(k)}) \right]^{-1}} \right].$$

A summary statistic of the $CPO_i's$ is the logarithm of the pseudo-marginal likelihood (LPML) defined as

$$LPML = \sum_{i=1}^{n} \log\left(CPO_i\right). \tag{2.51}$$

2.8 Application and Results

We have applied our models to welfare consumption, CPS size (positive), data using the nine covariates from NLSS-II. We have fitted models with the three standard distribution functions: the exponential, the gamma, and the generalized gamma distribution, assuming the response variable is noiseless. We fit our model without logarithmic transformation of the responses. We have sampled parameters using the MCMC Metropolis–Hastings algorithm along with a grid sampling method. For all the models fitted for noiseless responses, we have taken a set of 2100 samples, "burn-in" 100 samples and thinning interval of one. The final set has 1000 samples.

We have presented the acceptance rates for the final Metropolis–Hastings sampler, p-values of the Geweke convergence diagnostic tests, and effective sample sizes. For model comparison purpose, we have calculated LPML values. The larger the value of LPML the better the model. We have also presented the percentage of CPO values below 0.02 probability for each model.

Table 2.1 presents LPML values for models without random area effects and with random area effects for all six models developed in this chapter. For the models without random area effects, this table shows that the exponential models have much smaller values of LPML compared to gamma and generalized gamma models. So, obviously a gamma or generalized gamma model fits better for this data set. Comparing the gamma and generalized gamma models, the gamma model has larger LPML values than the generalized gamma model except in stratum 4.

For the models with random area effects, the LPML values in Table 2.1 show that the exponential model has very small LPML values compared with the gamma and generalized gamma models, which are also seen in the models without random area effects. The gamma model and generalized gamma model have very close LPML values, but the generalized gamma model shows bigger LPML than the gamma model for all strata except in stratum 2.

In Table 2.1 we have also provided a column for LPML values for hierarchical Bayesian nested error regression (NER) models. Since this hierarchical Bayesian NER model is built with a logarithmic transformation, to allow close comparison we exponentiate back the predicted responses and calculated the probability using the log-normal distribution. This table shows that a hierarchical Bayesian NER still has larger LPML than other models. But this model could be problematic under logarithmic transformation.

Table 2.2 presents the percentage of observations with CPO values below 0.02 probability for all models fitted: the exponential, the gamma, the generalized gamma, without random area effects, and with random area effects. Smaller percentage values are better. Here we see that every model without random area effects shows a percentage below 5%. Table 2.2 also presents the proportion of observations below 0.02 probability for models with random area effects. Here we see that there are some models which have values higher than 5%.

Table 2.3 presents the last Metropolis–Hastings sampler acceptance rate for parameters, the Geweke convergence diagnostic test, and effective sample sizes for the exponential model. The acceptance rate of parameters $\boldsymbol{\beta}$ for models without random area effects and parameters $(\boldsymbol{\beta}, \sigma^2)$ for models with random area effects are provided. The acceptance rates for $(\boldsymbol{\beta}, \sigma^2)$ are around 0.50 for all strata.

Table 2.4 presents the Metropolis–Hastings sampler acceptance rates for parameters, Geweke convergence diagnostic tests and effective sample sizes for the gamma model. The acceptance rate for parameters $(\alpha, \boldsymbol{\beta})$ for models without random area effects and parameters $(\alpha, \boldsymbol{\beta}, \sigma^2)$ for models with random area effects are provided. The acceptance rates for $(\alpha, \boldsymbol{\beta}, \sigma^2)$ are between 0.50 and 0.55 for all strata.

Table 2.5 presents the Metropolis–Hastings sampler acceptance rate for parameters, Geweke convergence diagnostic test, and effective sample sizes for the generalized gamma model. The acceptance rate for parameters $(\alpha, \gamma, \boldsymbol{\beta})$ of models without random area effects and parameters $(\alpha, \gamma, \boldsymbol{\beta}, \sigma^2)$ of models with random area effects are provided. The acceptance rates for $(\alpha, \gamma, \boldsymbol{\beta}, \sigma^2)$ are between 0.50 and 0.55 for all strata.

The trace and the correlation plots are from the generalized gamma model with random area effects. The trace and correlation plots are shown for the Mountains stratum (stratum one) as an example and all other strata have similar trace, density and correlation plots and not shown. The trace plots for parameters alpha, gamma, sigma square, and beta coefficients are shown from figure 2.1 to figure 2.13. The correlation plots for parameters alpha, gamma, sigma square, and vector of beta coefficients are shown from figure 2.14 to figure 2.26.

Below, we discuss the density plots of the responses and diagonal plots of the mean responses in the PSUs. The density plot of the observed responses with overlaying predicted responses, and the diagonal plot for observed mean responses versus predicted mean responses by PSUs are shown for all strata.

Figure 2.27 overlays the density plot of the observed welfare response variable and the density plots of 1000 responses predicted by the generalized gamma model with random area effects for the *Mountains stratum* (stratum one). The black line is for an observed response variable and red lines are for the predicted responses. Figure 2.28 shows the diagonal plot for comparing mean responses in the PSU for observed and predicted responses by the generalized gamma model with random area effects in the Mountains stratum.

Similarly, we present figures for all strata for generalized gamma models with random area effects. Figures 2.29 and 2.30 overlay density plots of the observed welfare response variable and predicted responses, and a diagonal plot for mean responses in the PSU for observed and predicted responses for stratum 2 *(Kathmandu valley urban areas).* Figures 2.31 and 2.32 overlay density plots of the observed welfare response variable and predicted responses, and a diagonal plot for mean responses in the PSU for observed and predicted responses for stratum 3 *(Other hills urban areas).* Figures 2.33 and 2.34 overlay density plots of observed welfare response variable and predicted responses, and a diagonal plot for mean responses in the PSU for observed and predicted responses for stratum 4 *(Hill rural areas).* Figures 2.35 and 2.36 overlay density plots of the observed welfare response variable and predicted responses, and a diagonal plot for mean responses in the PSU for observed and predicted responses for stratum 5 *(Terai urban areas)*. Figures 2.37 and 2.38 overlay density plots of the observed welfare response variable and predicted responses, and a diagonal plot for mean responses in the PSU for observed and predicted responses for stratum 6 *(Terai rural areas)*.

Finally, we note that, a formal article on noiseless CPS response data modeling without logarithmic transformation, is under preparation on the topic *"Hierarchical Bayesian models for size responses from small areas: An application to poverty estimation"* (Manandhar and Nandram, 2017a).

Table 2.1: LPML values for three standard models (with and without random area effects)

Models without random area effects

Stratum	**Model**		
	Expo	**Gamma**	**GGamma**
1	-434.1	-321.2	-332.4
2	-699.6	-643.7	-718.4
3	-485.9	-454.0	-475.2
4	-1344.7	-1098.0	-1073.3
5	-504.5	-464.7	-481.1
6	-1601.3	-1255.7	-1441.8

Models with random area effects

Stratum	**Model**			
	HB NER	**Expo**	**Gamma**	**GGamma**
1	-218.4	-559.5	-308.0	-223.1
2	-608.0	-922.1	-620.0	-626.7
3	-375.6	-884.3	-420.3	-417.3
4	-768.6	-2053.7	-1052.8	-949.5
5	-357.9	-894.4	-430.8	-424.5
6	-1022.0	-2195.7	-1248.4	-1178.2

Table 2.2: Percent of observations with CPO values below 0.02 for three models (with and without random area effects)

Models without random area effects

Stratum	Model		
	Expo	**Gamma**	**GGamma**
1	0.26	0.52	3.39
2	2.70	3.43	5.15
3	2.38	2.38	4.17
4	0.87	1.13	2.69
5	1.72	2.70	2.94
6	0.90	1.88	3.43

Models with random area effects

Stratum	Model			
	HB NER	**Expo**	**Gamma**	**GGamma**
1	2.08	1.56	0.26	0.78
2	3.68	8.33	3.19	3.68
3	2.38	13.10	2.08	2.08
4	1.82	5.03	0.61	1.39
5	2.94	11.52	1.72	1.96
6	2.53	4.41	1.31	1.88

Table 2.3: Exponential models: Metropolis–Hastings acceptance rates, Geweke test p-values and effective sample sizes

Model without area effects

Stratum	**Beta acceptance rate**	**P-values**									
		Beta0	**Beta1**	**Beta2**	**Beta3**	**Beta4**	**Beta5**	**Beta6**	**Beta7**	**Beta8**	**Beta9**
1	0.46	0.16	0.54	0.76	0.33	0.63	0.16	0.37	0.17	0.98	0.11
2	0.16	0.79	0.70	0.31	0.31	0.16	0.53	0.71	0.51	0.50	0.96
3	0.20	0.28	0.82	0.21	0.66	0.71	0.69	0.55	0.30	0.20	0.20
4	0.39	0.38	0.82	0.34	0.38	0.12	0.38	0.98	0.48	0.12	0.61
5	0.58	0.38	0.70	0.12	0.55	0.39	0.66	0.81	0.99	0.73	0.27
6	0.35	0.90	0.89	0.18	0.55	0.40	0.81	0.93	0.94	0.98	0.56

Stratum	**Effective sample size**									
	Beta0	**Beta1**	**Beta2**	**Beta3**	**Beta4**	**Beta5**	**Beta6**	**Beta7**	**Beta8**	**Beta9**
1	1.00	1.00	1.00	1.00	1.00	1.00	1.01	1.00	1.23	1.00
2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.88	0.90	1.00
3	0.89	1.01	1.00	1.14	1.00	1.00	1.00	1.00	0.59	1.00
4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
5	1.00	1.00	1.00	2.51	1.00	1.00	1.00	1.00	1.00	1.00
6	0.84	1.00	1.11	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Model with area effects

Stratum	**Sigma2 Beta acceptance rate**	**P-values**										
		Sigma	**Beta0**	**Beta1**	**Beta2**	**Beta3**	**Beta4**	**Beta5**	**Beta6**	**Beta7**	**Beta8**	**Beta9**
1	0.519	0.02	0.02	0.22	0.07	0.93	0.01	0.31	0.95	0.02	0.95	0.48
2	0.523	0.19	0.92	0.31	0.34	0.14	0.23	0.28	0.84	0.33	0.59	0.89
3	0.527	0.15	0.17	0.16	0.77	0.48	0.56	0.38	0.44	0.44	0.51	0.43
4	0.51	0.60	0.53	0.33	0.20	0.11	0.46	0.49	0.14	0.66	0.42	0.03
5	0.528	0.21	0.29	0.41	0.53	0.13	0.88	0.80	0.06	0.57	0.65	0.19
6	0.51	0.13	0.71	0.53	0.26	0.61	0.05	0.07	0.17	0.00	0.85	0.27

Stratum	**Effective sample size**										
	Sigma	**Beta0**	**Beta1**	**Beta2**	**Beta3**	**Beta4**	**Beta5**	**Beta6**	**Beta7**	**Beta8**	**Beta9**
1	1.00	1.00	1.00	0.85	0.77	1.00	1.00	1.00	1.00	1.00	0.88
2	1.00	1.22	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.27
3	1.00	1.00	1.00	1.00	1.22	1.13	1.00	1.01	0.83	0.84	1.00
4	1.00	1.00	1.00	1.14	1.00	1.05	0.73	1.11	1.00	1.00	1.00
5	1.00	0.91	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
6	0.89	1.00	1.00	1.00	1.00	1.00	1.00	0.87	1.00	1.00	1.00

Table 2.4: Gamma models: Metropolis–Hastings acceptance rates, Geweke test p-values and effective sample sizes

Model without area effects

Stratum	Alpha Beta acceptance rate	P-values										
		Alpha	Beta0	Beta1	Beta2	Beta3	Beta4	Beta5	Beta6	Beta7	Beta8	Beta9
1	0.439	0.97	0.53	0.91	0.62	0.10	0.58	0.41	0.20	0.57	0.83	0.65
2	0.552	1.00	0.68	0.48	0.86	0.54	0.95	0.72	0.65	0.72	0.46	0.98
3	0.537	0.52	0.29	0.37	0.72	0.61	0.57	0.62	0.46	0.76	0.09	0.36
4	0.345	0.26	0.11	0.41	0.59	0.53	0.56	0.68	0.46	0.51	0.46	0.40
5	0.516	0.50	0.81	0.77	0.83	0.05	0.71	0.47	0.45	0.30	0.21	0.51
6	0.312	0.40	0.29	0.46	0.42	0.36	0.35	0.09	0.50	0.64	0.53	0.31

Stratum	Effective sample size										
	Alpha	Beta0	Beta1	Beta2	Beta3	Beta4	Beta5	Beta6	Beta7	Beta8	Beta9
1	0.81	1.24	0.97	0.91	0.96	1.00	0.88	1.00	1.23	1.00	0.91
2	0.81	1.00	0.86	1.00	1.00	1.00	1.00	1.20	1.00	0.86	1.00
3	1.08	1.18	1.00	1.00	1.00	1.00	1.00	1.19	1.00	1.00	1.26
4	0.82	0.85	1.10	1.00	1.00	1.00	1.68	1.00	1.00	1.00	1.00
5	1.00	1.00	1.00	0.77	1.00	1.00	0.89	1.00	1.00	1.35	1.00
6	0.74	0.87	1.00	1.00	1.00	1.00	1.00	1.13	1.03	1.00	1.00

Model with area effects

Stratum	Alpha, Beta, Sigma2 acceptance rate	P-values											
		Alpha	Sigma2	Beta0	Beta1	Beta2	Beta3	Beta4	Beta5	Beta6	Beta7	Beta8	Beta9
1	0.529	0.45	0.66	0.58	0.40	0.43	0.92	0.42	0.66	0.50	0.72	0.78	0.21
2	0.523	0.40	0.96	0.40	0.52	0.12	0.99	0.93	0.25	0.50	0.66	0.39	0.36
3	0.525	0.86	0.31	0.95	0.52	0.60	0.02	0.79	0.78	0.96	0.74	0.94	0.78
4	0.511	0.74	0.53	0.99	0.65	0.27	0.33	0.06	0.68	1.00	0.07	0.23	0.07
5	0.524	0.13	0.58	0.98	0.04	0.45	0.10	0.88	0.05	0.50	0.19	0.24	0.08
6	0.512	0.54	0.28	0.93	0.40	0.52	0.98	0.62	0.83	0.31	0.78	0.20	0.68

Stratum	Effective sample size											
	Alpha	Sigma2	Beta0	Beta1	Beta2	Beta3	Beta4	Beta5	Beta6	Beta7	Beta8	Beta9
1	1.00	1.11	1.10	1.00	1.11	1.00	1.09	1.35	0.86	1.00	1.00	1.00
2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.92	1.00	1.00	0.81	1.00
3	1.00	1.00	1.00	0.94	1.09	1.00	1.00	1.00	0.82	1.00	1.00	2.25
4	1.15	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.79	1.00	0.98	1.00
5	1.00	1.00	1.00	1.00	1.00	1.00	0.80	1.00	1.00	1.11	1.12	1.00
6	1.00	0.80	1.00	1.11	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 2.5: Generalized Gamma models: Metropolis–Hastings acceptance rates, Geweke test p-values and effective sample sizes

Model without area effects

Stratum	**Alpha, Gamma, Beta acceptance rate**	**P-values**											
		Alpha	**Gamma**	**Beta0**	**Beta1**	**Beta2**	**Beta3**	**Beta4**	**Beta5**	**Beta6**	**Beta7**	**Beta8**	**Beta9**
1	0.505	0.45	0.28	0.21	0.33	0.21	0.94	0.36	0.80	0.95	0.19	0.07	0.87
2	0.502	0.55	0.61	0.37	0.26	0.79	0.83	0.48	0.92	0.70	0.71	0.70	0.25
3	0.497	0.67	0.66	0.55	0.12	0.96	0.47	0.08	0.62	0.13	0.13	0.53	0.21
4	0.502	1.00	0.94	0.94	0.87	0.68	0.39	0.19	0.20	0.78	0.53	0.27	0.34
5	0.516	0.73	0.60	0.20	0.18	0.20	0.19	0.11	0.03	0.30	0.48	0.84	0.21
6	0.503	0.74	0.72	0.75	0.37	0.10	0.55	0.32	0.52	0.61	0.94	0.56	0.74

Stratum	**Effective sample size**											
	Alpha	**Gamma**	**Beta0**	**Beta1**	**Beta2**	**Beta3**	**Beta4**	**Beta5**	**Beta6**	**Beta7**	**Beta8**	**Beta9**
1	1.00	1.00	1.00	0.78	0.91	1.26	0.89	1.00	0.89	1.00	1.00	0.90
2	1.17	1.00	1.11	1.00	1.00	0.90	1.00	1.00	1.00	1.00	1.22	1.09
3	1.28	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.91	1.00	0.96	1.00
4	1.27	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
5	1.00	0.91	1.00	1.00	1.00	1.00	1.11	1.00	1.13	1.43	0.86	1.00
6	1.00	1.00	1.00	1.13	1.00	1.00	1.00	1.00	1.09	1.00	1.00	1.12

Model with area effects

Stratum	**Alpha, Gamma Beta, Sigma2 acceptance rate**	**P-values**												
		Alpha	**Gamma**	**Sigma2**	**Beta0**	**Beta1**	**Beta2**	**Beta3**	**Beta4**	**Beta5**	**Beta6**	**Beta7**	**Beta8**	**Beta9**
1	0.521	0.53	0.70	0.54	0.17	0.62	0.34	0.43	0.61	0.37	0.48	0.18	0.40	0.22
2	0.512	0.37	0.33	0.43	0.18	0.39	0.24	0.06	0.55	0.56	0.50	0.38	0.16	0.13
3	0.515	0.01	0.27	0.97	0.01	0.02	0.06	0.65	0.79	0.05	0.09	0.06	0.20	0.42
4	0.507	0.51	0.29	0.42	0.42	0.80	0.93	0.48	0.53	0.43	0.87	0.29	0.25	0.61
5	0.517	0.59	0.19	0.58	0.76	0.09	0.37	0.38	0.73	0.67	0.44	0.20	0.03	0.32
6	0.512	0.43	0.96	0.47	0.47	0.22	0.35	0.34	0.60	0.19	0.54	0.53	0.70	0.57

Stratum	**Effective sample size**												
	Alpha	**Gamma**	**Sigma2**	**Beta0**	**Beta1**	**Beta2**	**Beta3**	**Beta4**	**Beta5**	**Beta6**	**Beta7**	**Beta8**	**Beta9**
1	1.10	1.10	1.00	1.00	1.09	1.00	1.00	1.00	1.00	1.11	1.00	1.00	1.00
2	1.14	1.18	1.00	1.54	1.00	1.00	1.10	1.00	1.00	1.11	1.16	1.31	1.24
3	0.83	0.89	1.10	0.91	1.00	0.91	1.00	1.00	0.78	0.89	0.90	0.90	1.00
4	1.00	1.00	1.00	1.00	1.00	1.00	1.16	1.00	0.81	1.06	1.00	1.00	1.00
5	1.01	1.11	1.00	1.01	1.00	1.00	1.00	1.00	1.12	1.00	0.94	1.00	1.00
6	1.26	1.30	1.00	1.30	1.00	1.18	1.28	1.00	1.00	1.00	1.34	1.19	1.25

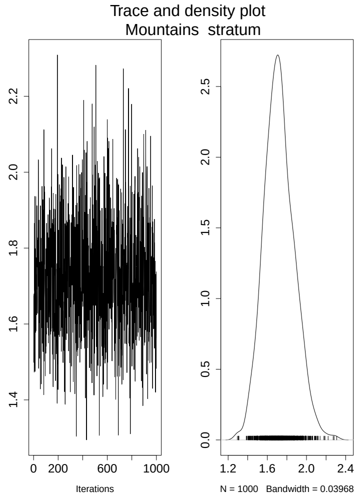


Figure 2.1: Trace plot: Alpha

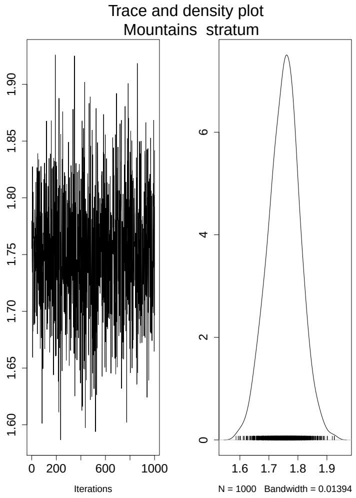


Figure 2.2: Gamma

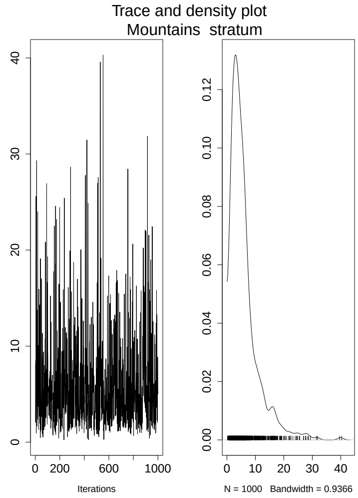


Figure 2.3: Sigma Square

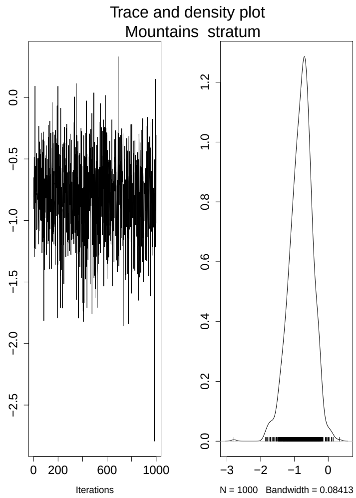


Figure 2.4: Beta0

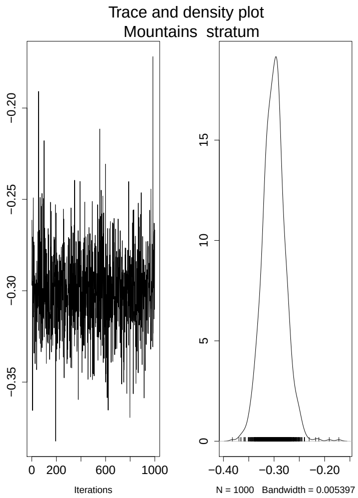


Figure 2.5: Beta1

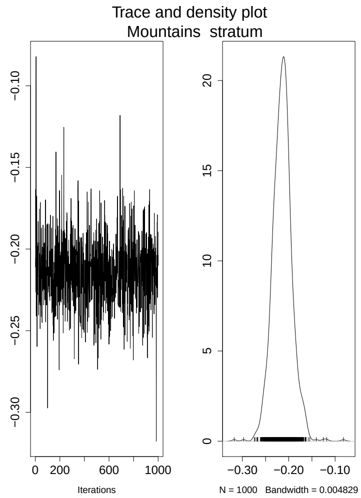


Figure 2.6: Beta2

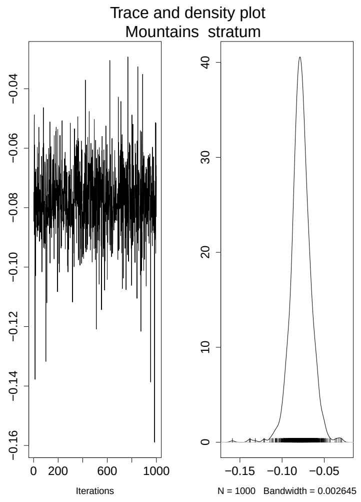


Figure 2.7: Beta3

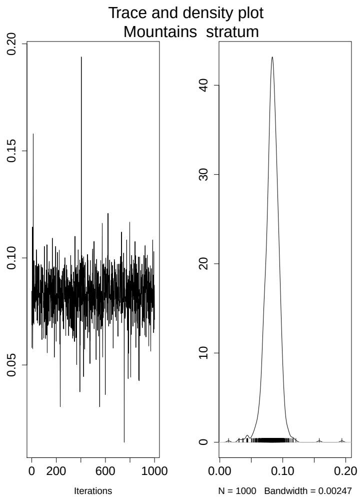


Figure 2.8: Beta4

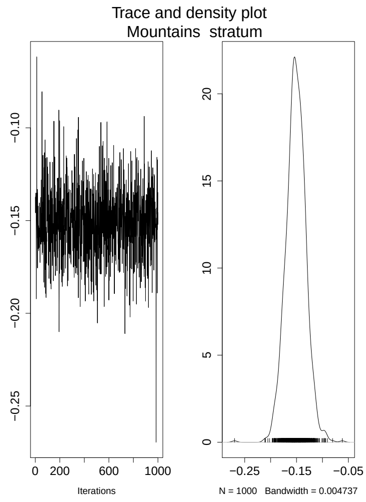


Figure 2.9: Beta5

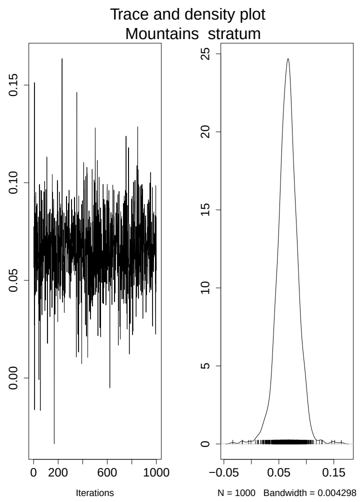


Figure 2.10: Beta6

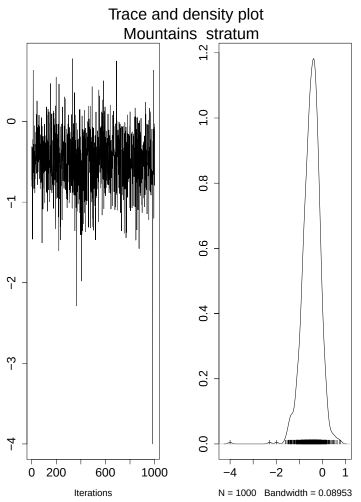


Figure 2.11: Beta7

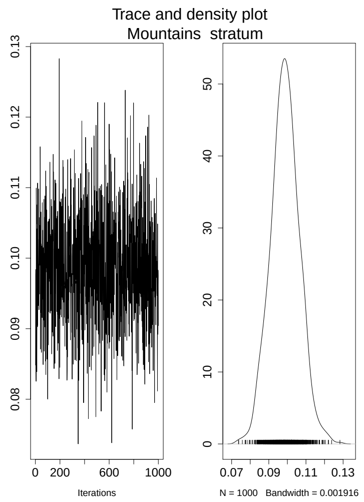


Figure 2.12: Beta8

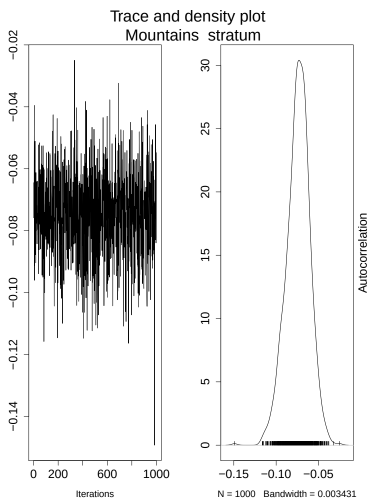


Figure 2.13: Beta9

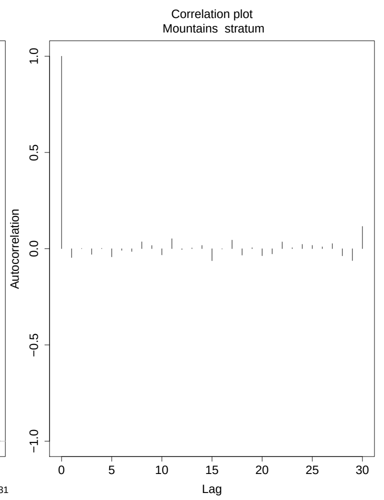


Figure 2.14: Alpha

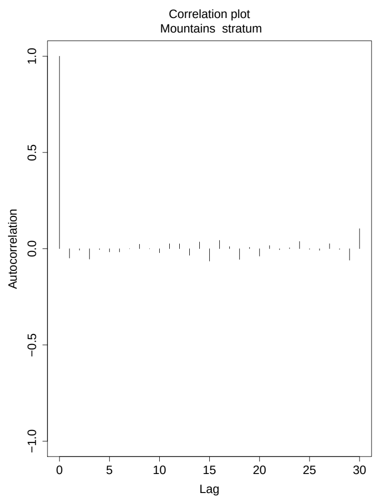


Figure 2.15: Gamma

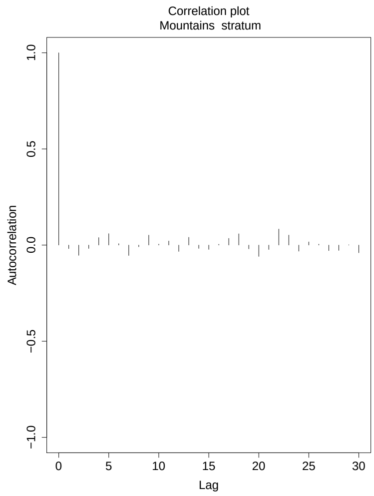


Figure 2.16: Sigma Square

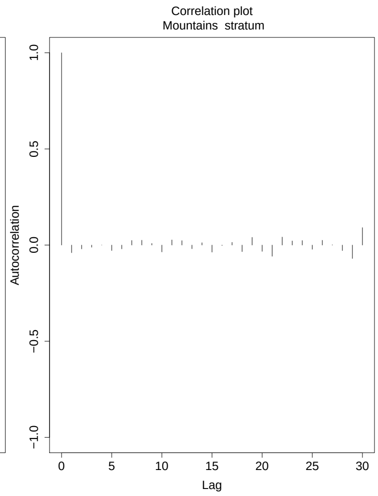


Figure 2.17: Beta0

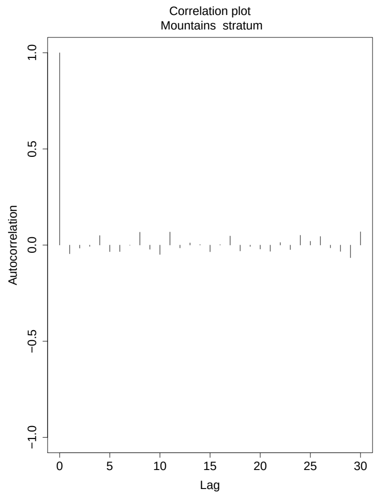


Figure 2.18: Beta1

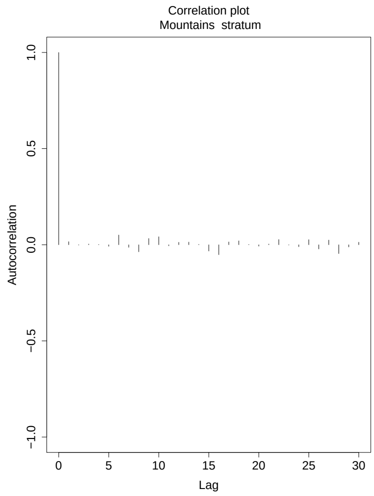


Figure 2.19: Beta2

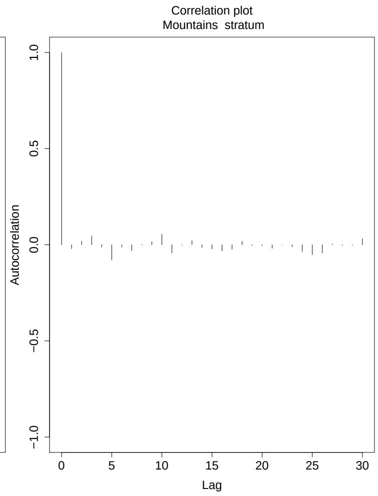


Figure 2.20: Beta3

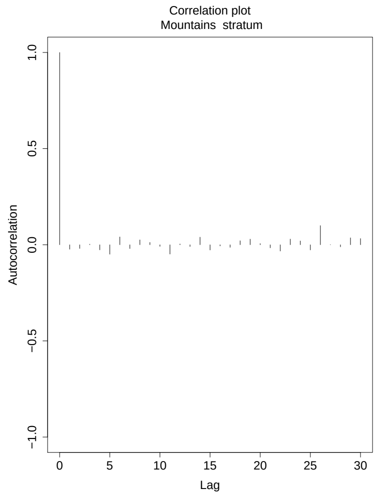


Figure 2.21: Beta4

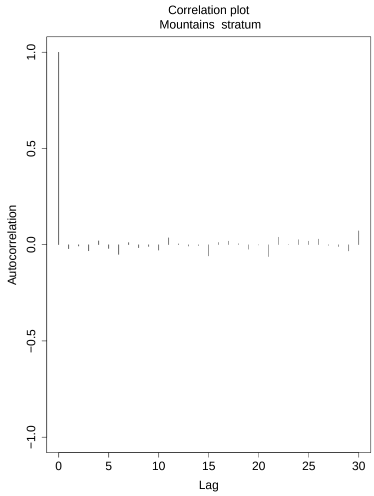


Figure 2.22: Beta5

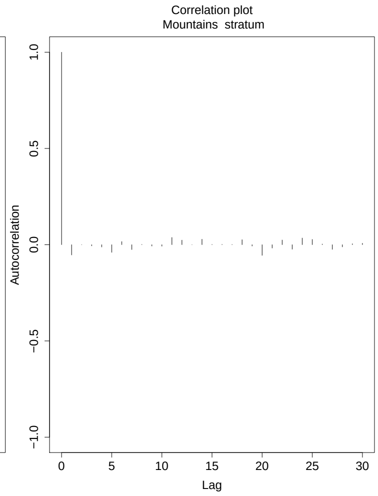


Figure 2.23: Beta6

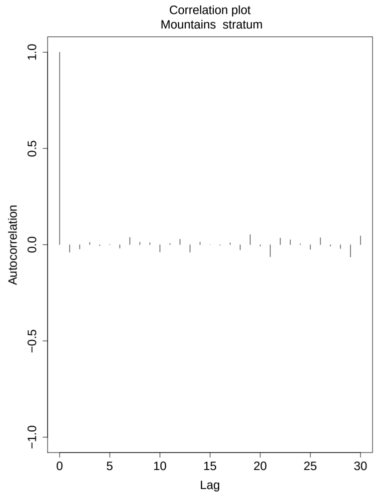


Figure 2.24: Beta7

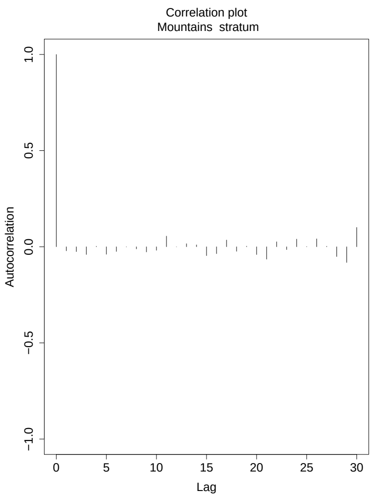


Figure 2.25: Beta8

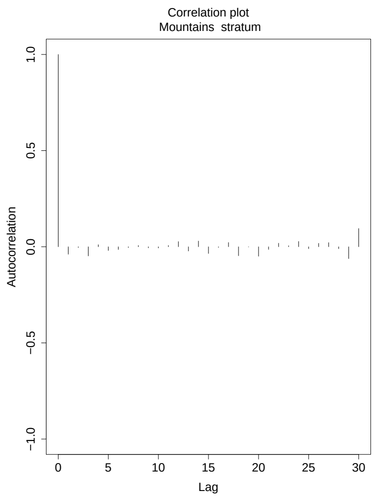


Figure 2.26: Beta9

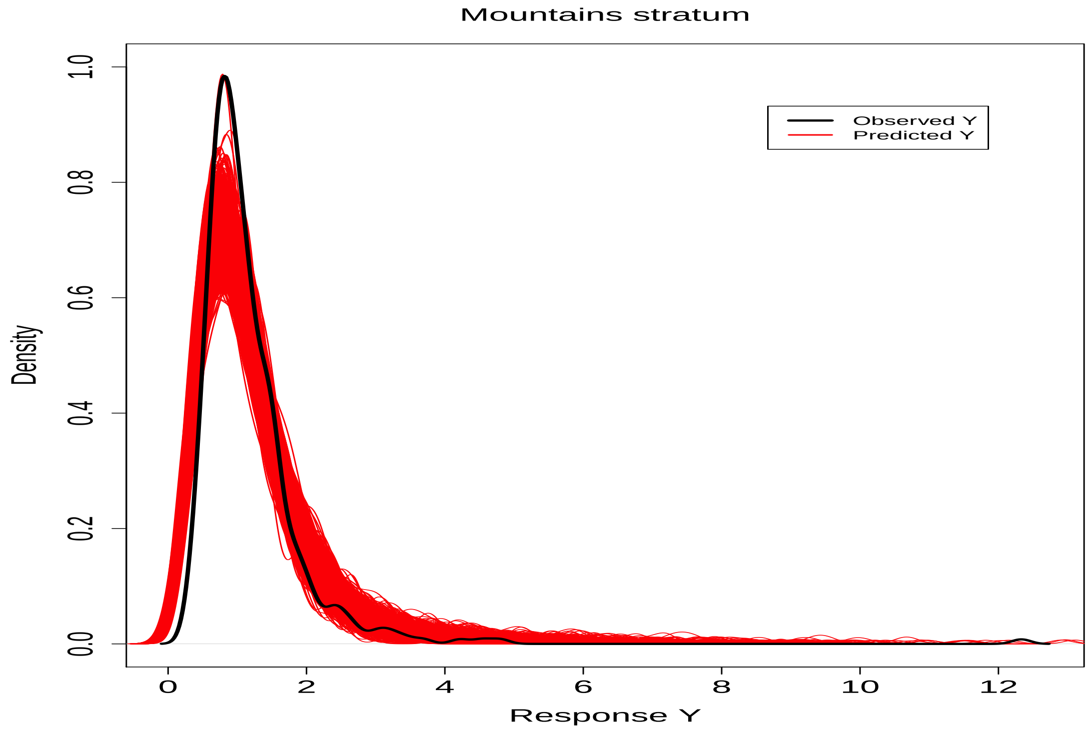


Figure 2.27: Stratum 1: density plots of observed and Generalized gamma model with random area effects' predicted responses

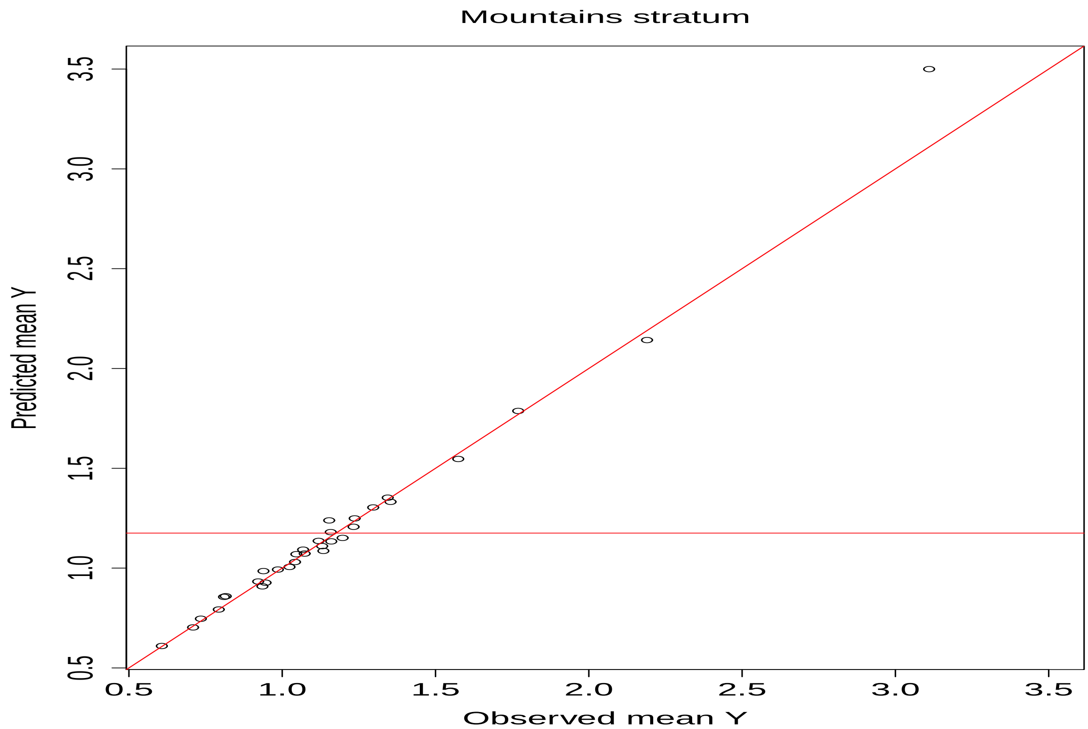


Figure 2.28: Stratum 1: Observed and Generalized gamma model with random area effects' predicted mean responses by PSU

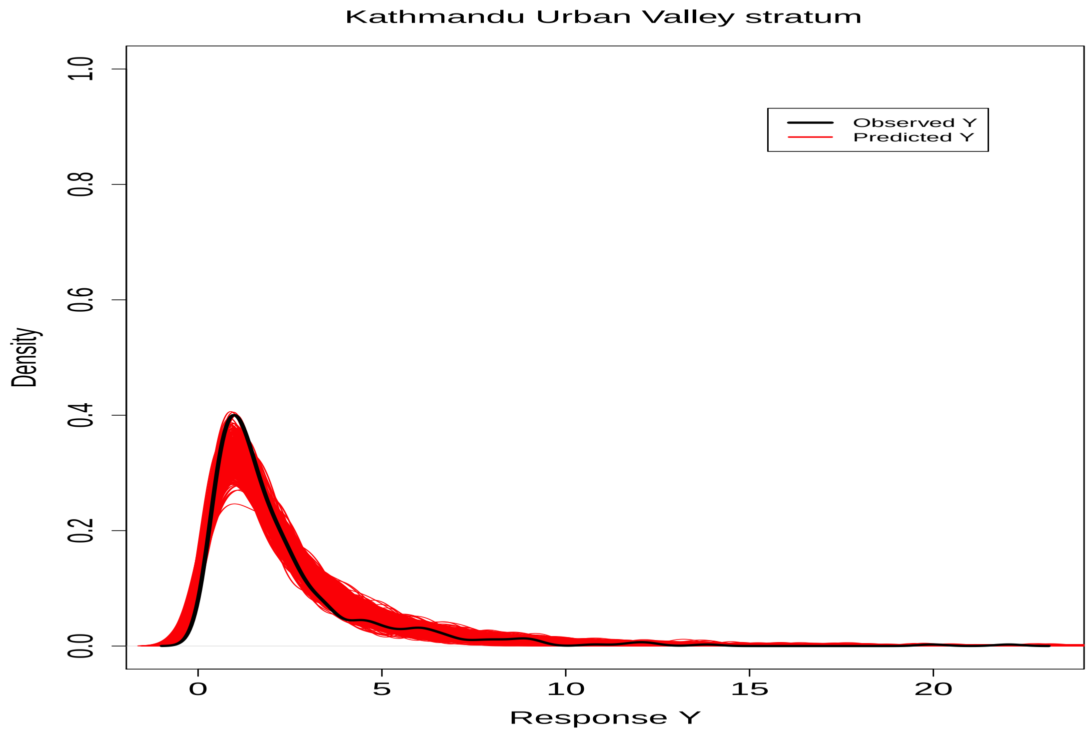


Figure 2.29: Observed and predicted responses density plots

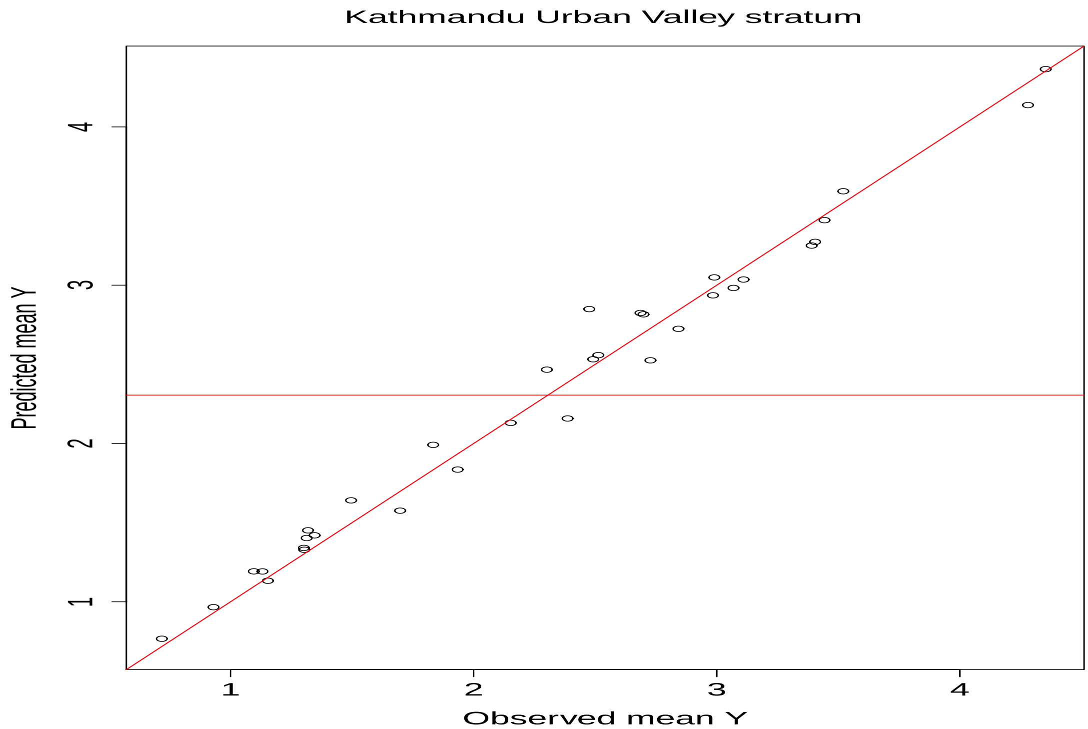


Figure 2.30: Observed and predicted mean responses by PSU

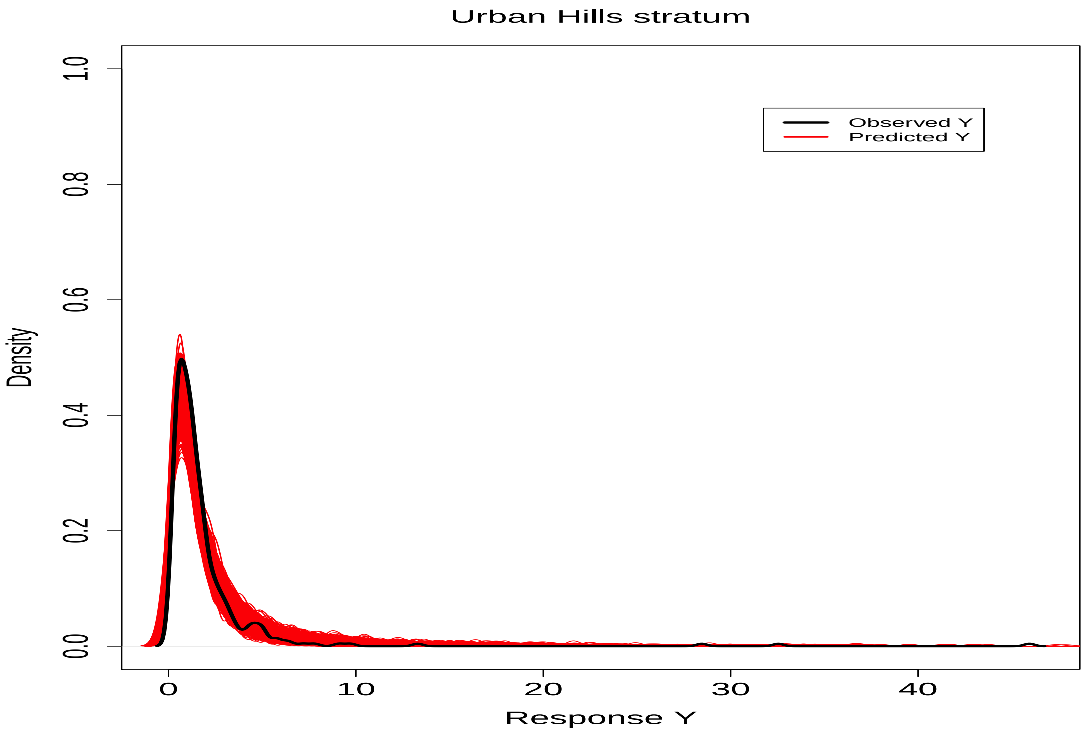


Figure 2.31: Observed and predicted responses density plots

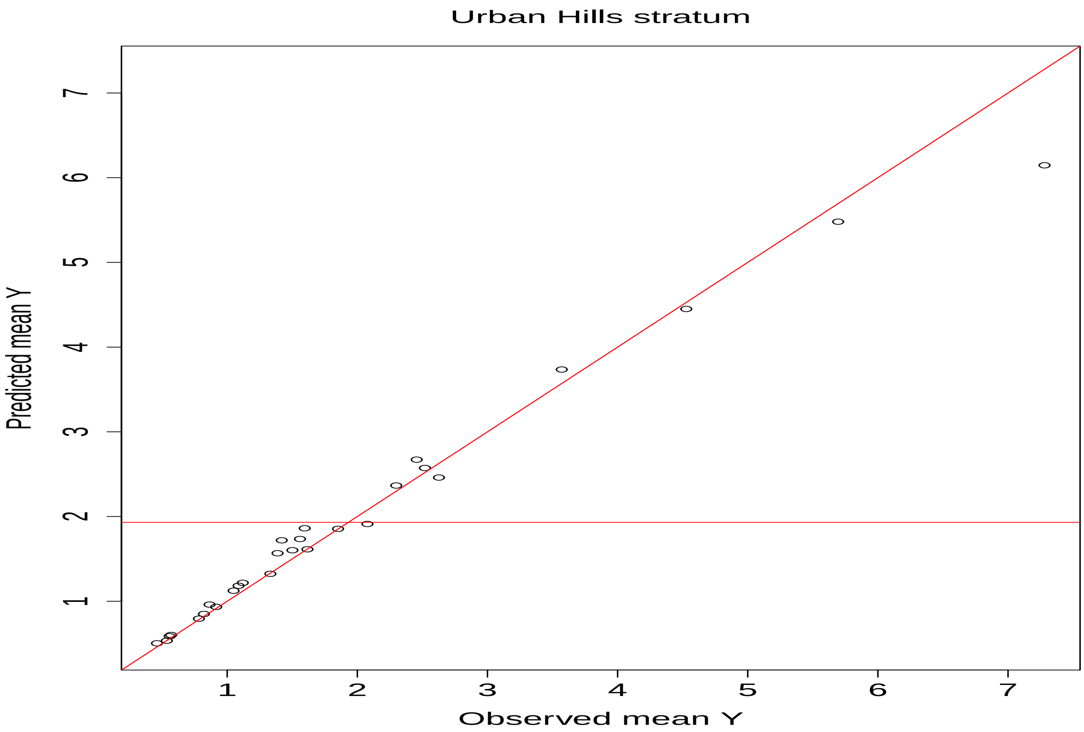


Figure 2.32: Observed and predicted mean responses by PSU

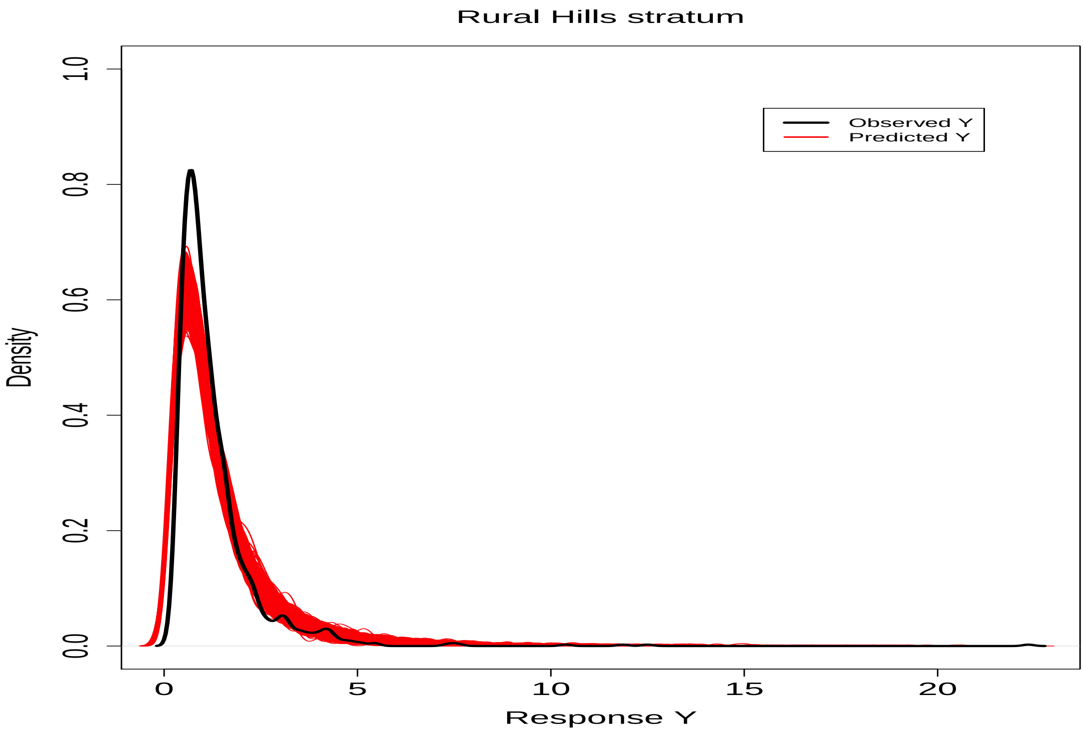


Figure 2.33: Observed and predicted responses density plots

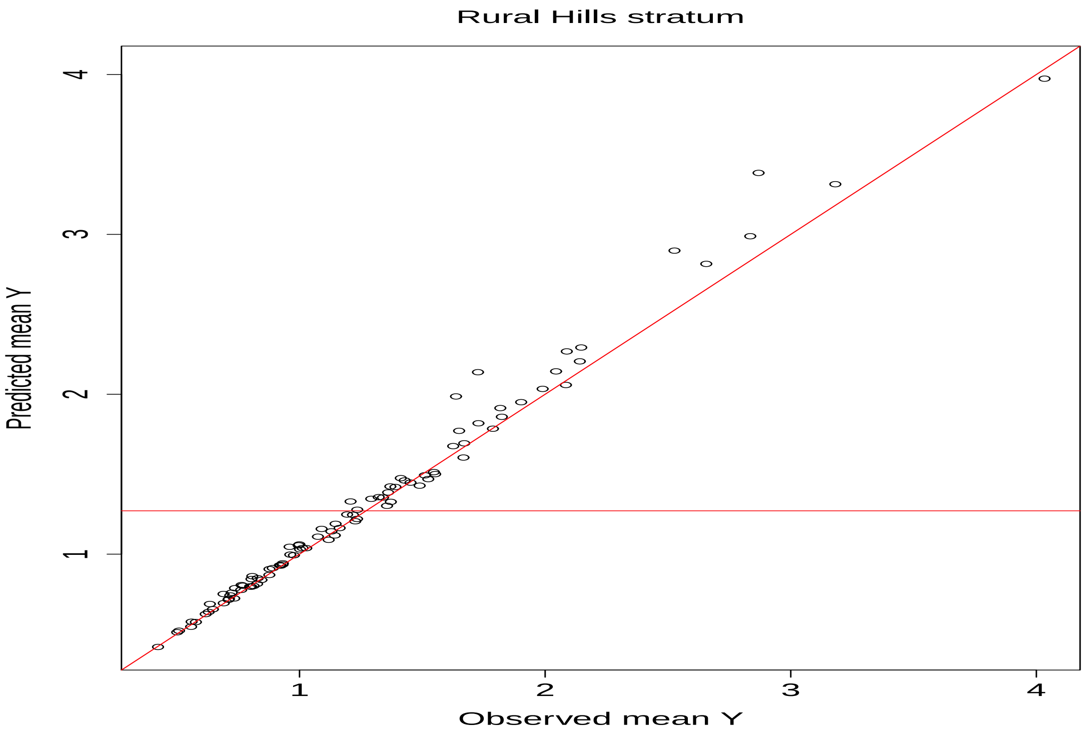


Figure 2.34: Observed and predicted mean responses by PSU

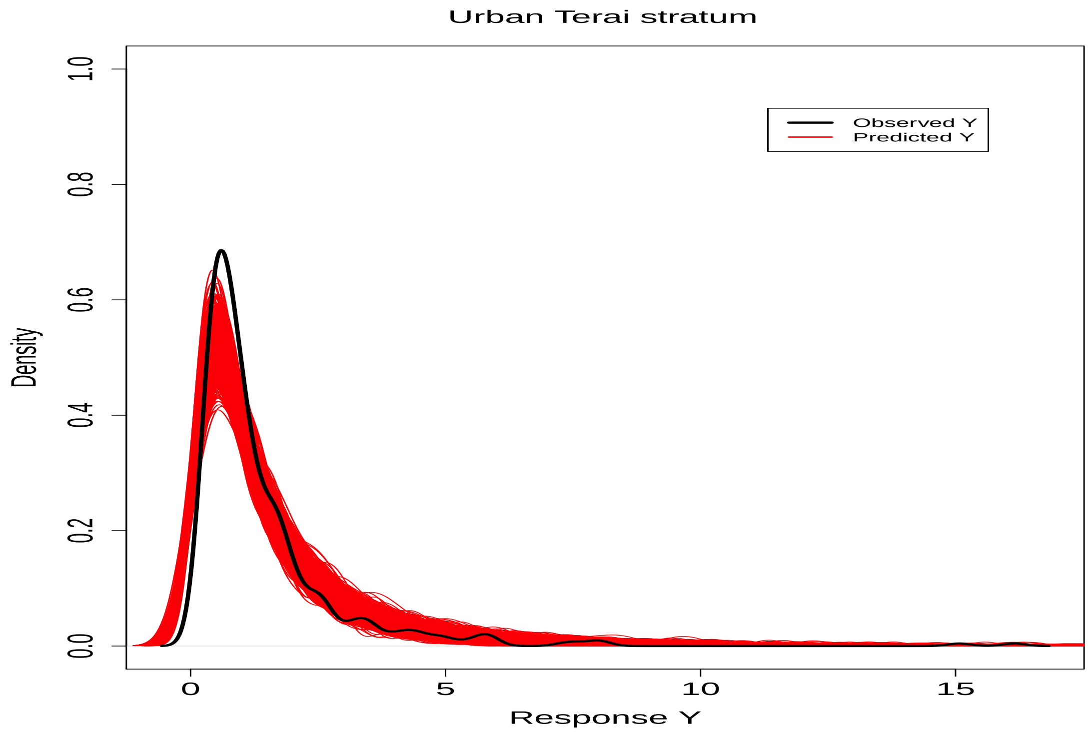


Figure 2.35: Observed and predicted responses density plots

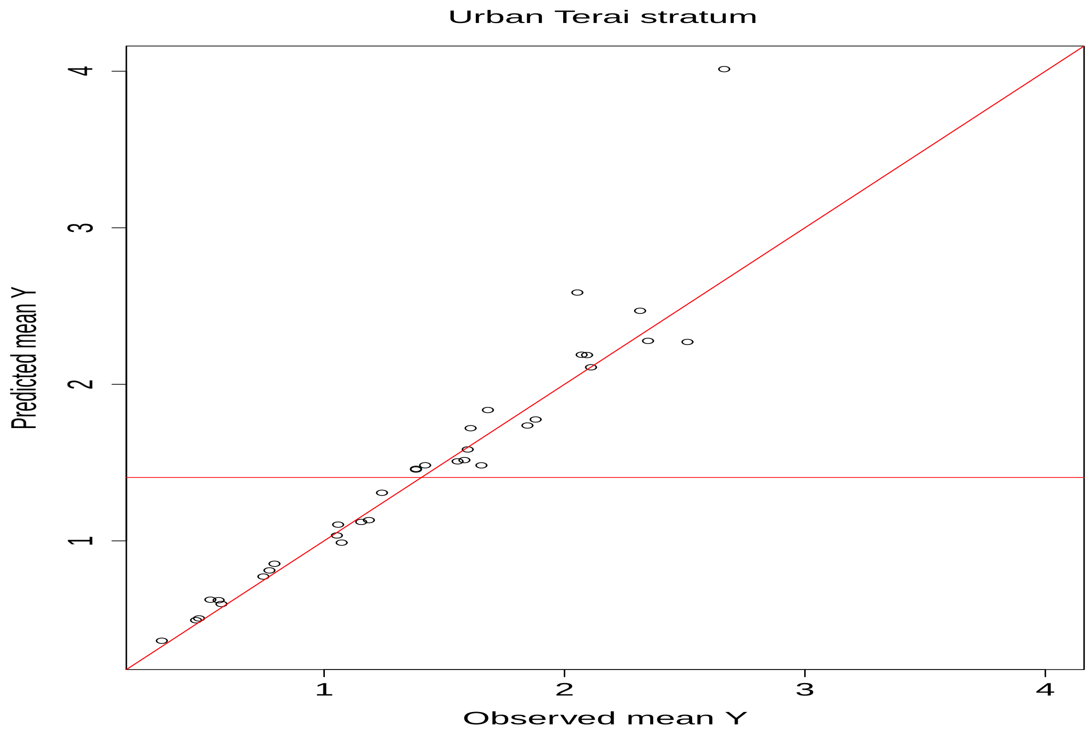


Figure 2.36: Observed and predicted mean responses by PSU

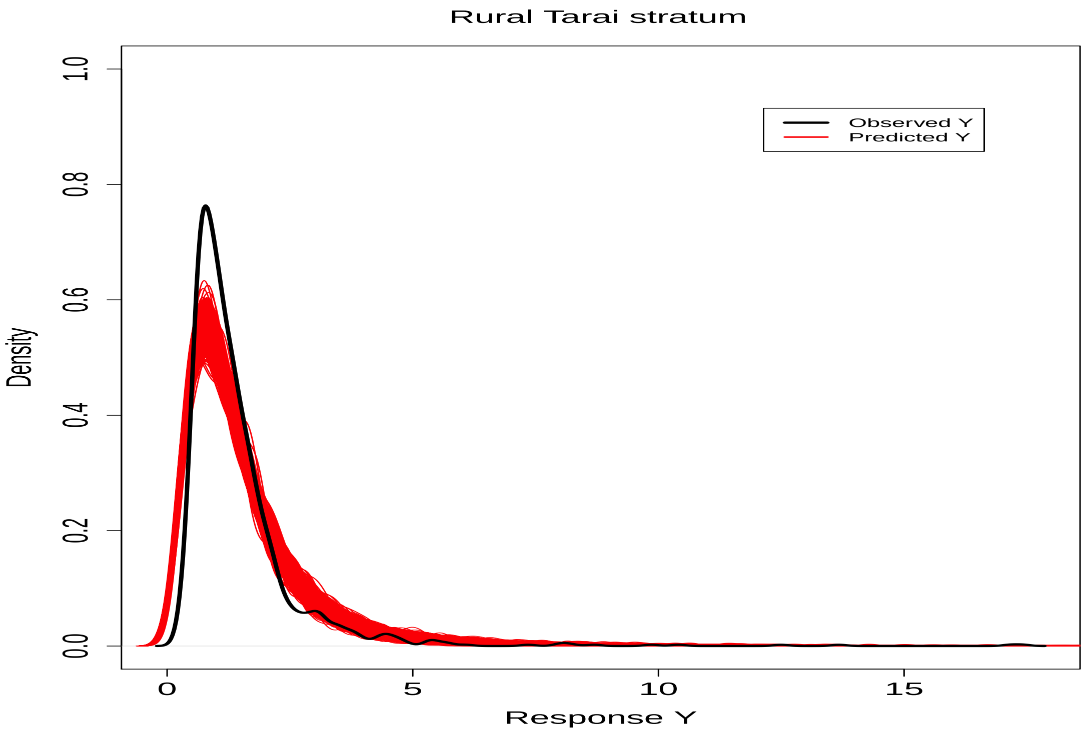


Figure 2.37: Observed and predicted responses density plots

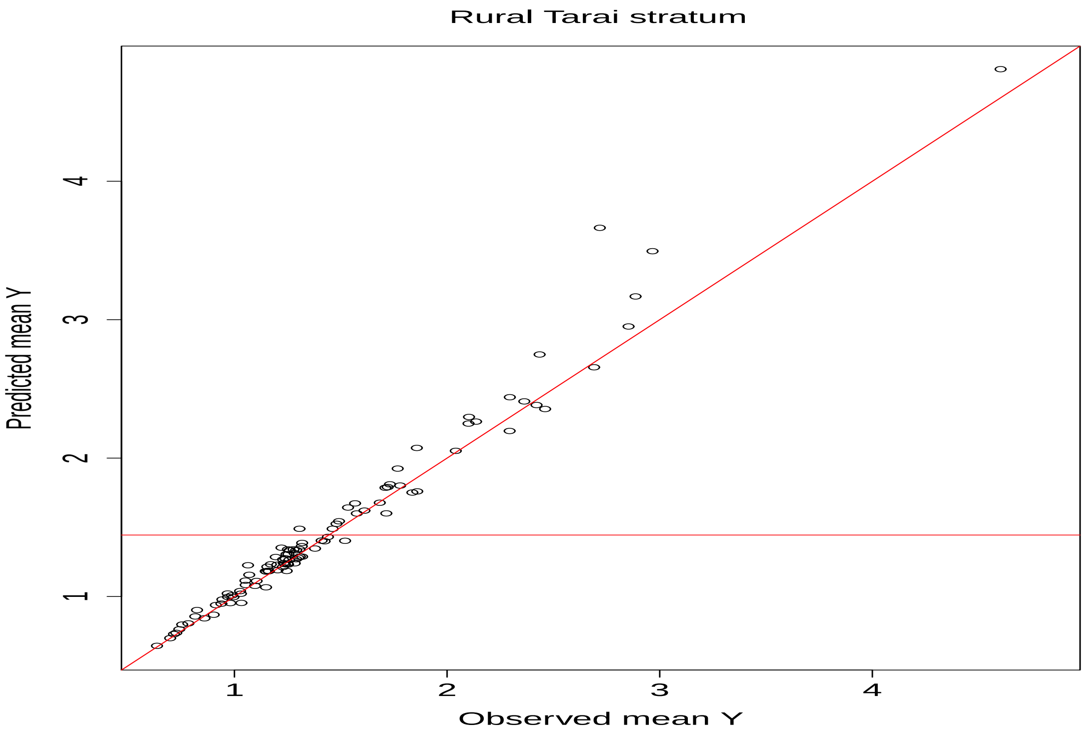


Figure 2.38: Observed and predicted mean responses by PSU

Chapter 3

Models for Noisy Responses

In this chapter we assume that the continuous and positively-skewed (CPS) responses are noisy. We define noise as response errors, recall errors, bias, or other errors introduced in the response data. See also Chapter 1, subsection *"Response per capita consumption"*. For the responses with noise, we fit a flexible distribution, a mixture of generalized gamma GB2 distribution, rather than fitting the standard distributions as in Chapter 2.

As discussed in Chapter 1, subsection *"Modeling with GB2 distribution"*, the GB2 distribution has four parameters which can be expressed as a mixture of the two generalized gamma distributions. The probability density function of each response variable $y|\alpha, \lambda, \gamma$ and the probability density function of its rate parameter $\lambda|\phi, \theta, \gamma$ both having the generalized gamma distribution. Mixing these two generalized gamma distributions gives a $GB2$ distribution with four parameters $Y \sim GB2(\alpha, \phi, \gamma, \theta)$. In the GB2, the shape parameters α and ϕ determine the skewness of the distribution, the shape parameter γ controls the overall shapes, and θ is the rate parameter. In GB2 we do not observe the rate parameter λ of the response variable directly, which has another generalized-gamma distribution, and we exploit this phenomenon to describe the noisy responses. So, in the GB2 distribution the rate parameter of the response variable $\boldsymbol{y}$ is hidden, which has linked with the shape and rate parameters of its own generalized gamma distribution. Therefore, the GB2 distribution has one more fold of distribution than the standard distributions we discussed in Chapter 2. We introduce the covariates in GB2 models through θ, the rate parameter of λ distribution.

As we know, if the random variable $Y|\alpha, \lambda, \gamma$ has the generalized gamma distribution with the shape parameters α, γ and the rate parameter λ, then it has the gamma distribution when γ equals one, and it has the exponential distribution when α and γ both equal one. Since GB2 is a mixture of two generalized gamma distributions, we explore some possible mixtures. In this chapter, we consider the three special cases of GB2 distributions: (i) the mixture of the exponential and the gamma distributions, (ii) the mixture of the two gamma distributions, and (iii) the mixture of the two generalized gamma distributions. For each GB2 distribution we have chosen, we fit two models, one without random area effects and another with random area effects.

In our model, the joint posterior density and the conditional posterior density functions are not in simple forms, so we use the second-order Taylor's series approximation to the unimodal GB2 density function to facilitate the sampling procedures. Taylor's series approximation helps us to approximate the complex density by multivariate normal. We have sampled the parameters using the grid sampling method and the Metropolis–Hastings (MH) algorithm. We have multivariate t-distributions as the proposal distributions with $d = 3$ degrees of freedom for the MH algorithm. We take 3 degrees of freedom so that variance will exist.

We have CPS welfare consumption response, assumed to be noisy, with nine covariates. This is the same data set as we used in Chapter 2 for modeling noiseless responses. The covariates are (i) "Household size" (*hhsize*), (ii) "proportion of kids aged 0 - 6 in the household" (*skids6*), (iii) "proportion of kids aged 7 - 14 in the household" (*skids714*), (iv) "abroad migrant" (*remtab*), (v) "House temporary" (*hutype3*), (vi) "House owned" (*huown2*), (vii) "proportion of households with cooking fuel LP/gas in Ward" (*ckfuel3w*), (viii) "proportion of household with land-owning females in municipality/VDC" (*pflandv*), and (ix) "proportion of kids 6-16 attending school in municipality/VDC" (*pschv*) from NLSS-II 2003–2004. In this chapter we assume that the noisy responses are better modeled by a flexible GB2 distribution rather than a standard distribution. We fit the GB2 distribution for noisy responses without logarithmic transformation of the response variable, see also Chapter 1, subsection *"Modeling with GB2 distribution."*

We consider three special cases of GB2: mixture of the exponential and the gamma, mixture of the two gamma, and the mixture of the two generalized gamma distributions. For each special case, we fit two hierarchical Bayesian models, one without random area effects and another with random area effects. We calculate CPO and find summary statistics LPML for model comparisons.

Notation: Let the distribution of the response variable and its rate parameter have the generalized gamma distribution

$$y|\alpha, \lambda, \gamma \sim \text{GGamma}(\alpha, \lambda, \gamma),$$
$$\lambda|\phi, \theta, \gamma \sim \text{GGamma}(\phi, \theta, \gamma).$$

Consider sample data with n observations, response variable $\boldsymbol{y}_{n\times 1}$, and covariate $\boldsymbol{x}_{n\times p}$. Here, the i^{th} observed response y_i has its corresponding covariate $\boldsymbol{x}_i,\ i = 1, \cdots, n$. To build models without random area effects, we introduce covariates through the rate parameter θ as $e^{\boldsymbol{x}_i'\boldsymbol{\beta}}$.

To build models with random area effects, we have ℓ small areas, $i = 1, \cdots, \ell$, and each small area has $j = 1, \cdots, n_i$, observations. Let y_{ij} and $\boldsymbol{x}_{ij},\ i = 1 \cdots, \ell,\ j = 1 \cdots, n_i$ denote the response variable and the corresponding covariates in the i^{th} area and j^{th} observation. We introduce covariates through the rate parameter θ as $e^{\boldsymbol{x'}_{ij}\boldsymbol{\beta}+\nu_i}$, where ν_i is the random area effect for the i^{th} area.

3.1 Exponential-Gamma Mixture GB2 Model Without Random Area Effects

The simplest mixture for GB2 density is the mixture of two exponential distributions. However, it does not exist which we have already discussed the mixture of the two exponential GB2 distributions in Chapter 1, subsection *"Modeling with GB2 Distribution"* paragraph *"Exponential-Gamma Mixture Model"*. Therefore, the simplest GB2 model that we can have is the mixture of the exponential and the gamma distribution. Let the response variable have the exponential distribution and its rate parameter follows the gamma distribution,

$$
\begin{aligned}
f(y|\lambda) &= \lambda e^{-\lambda y}, \quad \lambda > 0, \\
f(\lambda|\alpha, \theta) &= \frac{e^{-\theta y}\,\lambda^{\alpha-1}}{\Gamma(\alpha)}\,\theta^{\alpha}, \quad \alpha, \theta > 0.
\end{aligned}
$$

Mixing these two distributions and integrating out rate parameter λ, we get the GB2 density, the exponential-gamma mixture as

$$
\begin{aligned}
f(y|\alpha, \theta) &= \int_{\lambda} \lambda e^{-\lambda y}\,\frac{e^{-\theta y}\,y^{\alpha-1}}{\Gamma(\alpha)}\,\theta^{\alpha}\,d\lambda \\
&= \frac{\alpha}{\theta(1+\frac{y}{\theta})^{\alpha+1}}, \quad \alpha, \theta > 0.
\end{aligned}
$$

The k^{th} moment of the response is given by

$$
E[Y^k|\alpha, \theta] = \Gamma(k+1)\,\frac{\Gamma(\alpha-k)}{\Gamma(\alpha)}\,\theta^k, \quad \alpha > k. \tag{3.1}
$$

It shows from (3.1) that, we need $\alpha > 2$ for the variance to exist. We note that rate parameter α and θ are not identifiable. So we keep a restriction of $\alpha = 3$ in our model. Here, we consider the distribution of the response variable $Y|\lambda$ as the gamma distribution with shape $\alpha = 1$ and the distribution of $\lambda|\alpha, \theta$ is the gamma distribution with shape $\alpha = 2$ and assumed addition of the same shape parameter.

We assume that the responses $y_i|\alpha, \theta$, $i = 1, \cdots, n$ are random samples from the GB2 distribution, the mixture of the exponential and the gamma distributions, with rate $e^{\boldsymbol{x}_i'\boldsymbol{\beta}}$ and the shape parameter α fixed at 3. For simplicity, we write only GB2 afterward in this section to denote the GB2 distribution as a mixture of the exponential and the gamma distributions as defined above with $\alpha = 3$ fixed. The likelihood function is given by

$$
\pi(y_i|\alpha = 3, \boldsymbol{\beta}) = \prod_{i=1}^{n} \frac{\alpha\, e^{-\boldsymbol{x}_i'\boldsymbol{\beta}}}{\left(1 + y_i\, e^{-\boldsymbol{x}_i'\boldsymbol{\beta}}\right)^{\alpha+1}}.
$$

To build a Bayesian model we consider a non-informative prior for $\boldsymbol{\beta}$. The GB2 model without random area effects is

$$
\begin{aligned}
&y_i|\alpha = 3, \boldsymbol{\beta} \overset{\text{ind}}{\sim} \mathrm{GB2}(\alpha = 3, e^{\boldsymbol{x}_i'\boldsymbol{\beta}}), \quad \theta_i = e^{\boldsymbol{x}_i'\boldsymbol{\beta}},\ i = 1, \cdots, n, \\
&\pi(\boldsymbol{\beta}) \propto 1.
\end{aligned} \tag{3.2}
$$

The posterior density function is

$$\begin{aligned} \pi(\boldsymbol{\beta}|\boldsymbol{y}) &\propto f(\boldsymbol{y}|\alpha=3,\boldsymbol{\beta})\,\pi(\boldsymbol{\beta}) \\ &= \frac{e^{-\sum_{i=1}^{n}\boldsymbol{x_i'\beta}}}{\prod_{i=1}^{n}\left(1+y_i\,e^{-\boldsymbol{x_i'\beta}}\right)^{\alpha+1}}. \end{aligned} \tag{3.3}$$

Let the log-likelihood function be $G(\boldsymbol{\beta}|\boldsymbol{y}) = log(f(\boldsymbol{y}|\boldsymbol{\beta}))$,

$$G(\boldsymbol{\beta}|\boldsymbol{y}) = n\alpha - \sum_{i=1}^{n}\boldsymbol{x_i'\beta}\; - \;(\alpha+1)\sum_{i=1}^{n}\log(1+y_i\,e^{-\boldsymbol{x_i'\beta}}).$$

For notational simplicity, let us write G for the log-likelihood function, then its first- and second-order partial derivatives with respect to $\boldsymbol{\beta}$ are given by

$$\begin{aligned} \frac{\partial G}{\partial\boldsymbol{\beta}} &= -\sum_{i=1}^{n}\left[1\,+\,(\alpha-1)\,\left(1+\frac{e^{-\boldsymbol{x_i'\beta}}}{y_i}\right)^{-1}\right]\boldsymbol{x_i}, \\ \frac{\partial^2 G}{\partial\boldsymbol{\beta}^2} &= -\,(\alpha+1)\sum_{i=1}^{n}\frac{y_i\,e^{-\boldsymbol{x_i'\beta}}}{\left(1+y_i\,e^{-\boldsymbol{x_i'\beta}}\right)^2}\,\boldsymbol{x_i x_i'}. \end{aligned}$$

Using the first-order Taylor's series approximation for $\frac{y_i\,e^{-\boldsymbol{x_i'\beta}}}{\left(1+y_i\,e^{-\boldsymbol{x_i'\beta}}\right)^2}$ at $\boldsymbol{\beta}=\boldsymbol{0}$, the approximate MLE of $\boldsymbol{\beta}$ is given by

$$\boldsymbol{\beta}^*|\alpha = \left[\sum_{i=1}^{n}\frac{y_i}{(1+y_i)^2}\boldsymbol{x_i x_i'}\right]^{-1}\left(\sum_{i=1}^{n}\left(\frac{y_i}{1+y_i}-\frac{1}{\alpha+1}\right)\boldsymbol{x_i}\right). \tag{3.4}$$

Let the gradient vectors and the Hessian matrix evaluated at the approximate mode values $\boldsymbol{\beta}^*$ be $\nabla G(\beta^*)$ and $H(\beta^*)$ respectively. Using the *multivariate normal approximation theorem* from Chapter 1 we can write the approximated likelihood density function as

$$f(\boldsymbol{\beta}|\alpha,\boldsymbol{y}) \propto N_{\boldsymbol{\beta}}\left[\beta^* + (-H(\boldsymbol{\beta^*}))^{-1}\,\nabla G(\boldsymbol{\beta^*}),\,(-H(\boldsymbol{\beta^*}))^{-1}\right].$$

From the above distribution it follows that $\boldsymbol{\beta}$ has the multivariate normal distribution given by

$$\boldsymbol{\beta}|\alpha,\boldsymbol{y} \sim \mathrm{MN}\left(\beta^* + (-H(\boldsymbol{\beta^*}))^{-1}\,\nabla G(\boldsymbol{\beta^*}),\,(-H(\boldsymbol{\beta^*}))^{-1}\right). \tag{3.5}$$

3.1.1 Sampling from Joint Posterior Density

We have fixed $\alpha = 3$ in this GB2 model and the only unknown parameter is $\boldsymbol{\beta}$. We draw $\boldsymbol{\beta}$ using the Metropolis–Hastings algorithm. The proposal distribution for $\boldsymbol{\beta}$ is the multivariate t-distribution with d degrees of freedom. The mean and covariance matrix for the proposal distribution are obtained from 1,000 samples drawn from (3.5). The target distribution is the posterior distribution function (3.3). In the MCMC sequence we keep samples only if it moves. We test the convergence of the samples and checked the acceptance rate of the samples. We draw a set of 1,000 samples.

3.1.2 Prediction

After drawing a set of $\boldsymbol{\beta}$ parameters from the GB2 model, we predict response variables as follows:

(i) Find the rate parameters

$$\theta_i = e^{\boldsymbol{x}_i'\boldsymbol{\beta}}.$$

(ii) Find the rate parameters λ_i from the gamma distribution

$$\lambda_i \sim \text{Gamma}(\alpha = 3, \theta_i).$$

(iii) Predict the responses from the exponential distribution

$$\hat{y}_i \sim \text{Expo}(\lambda_i).$$

3.2 Exponential-Gamma Mixture GB2 Model With Random Area Effects

In this section we discuss the simplest GB2 model with random area effects, the mixture of the exponential and the gamma distributions. As in section 3.1, We have $Y|\lambda \sim \text{Expo}(\lambda)$ and its rate parameter $\lambda|\alpha, \theta \sim \text{Gamma}(\alpha, \theta)$. We fix $\alpha = 3$ as discussed in previous section 3.1, so that the variance will exist and parameters are estimable. For simplicity, we write GB2 afterward in this section to indicate the GB2 distribution as a mixture of the exponential and the gamma distribution with random area effects $\boldsymbol{\nu}$. We assume that the responses $y_{ij}|\alpha, \theta, i = 1, \cdots, \ell,\ j = 1, \cdots, n_i$ are independent random samples from the GB2 distribution and ν_i follows the normal distribution with mean zero and the variance σ^2. We introduce covariates in the model through the rate parameter θ as $e^{\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\boldsymbol{\nu_i}}$. The likelihood function is

$$\pi(\boldsymbol{y}|\alpha = 3, \boldsymbol{\beta}) = \prod_{i=1}^{\ell}\prod_{j=1}^{n_i} \frac{\alpha\, e^{-(\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\nu_i)}}{\left(1 + y_{ij}\, e^{-(\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\nu_i)}\right)^{\alpha+1}}. \tag{3.6}$$

Let $\boldsymbol{\beta}$ and σ^2 have non-informative independent priors. The hierarchical Bayesian GB2 model with random area effects is

$$\begin{aligned} y_{ij}|\boldsymbol{\beta}, \nu_i &\overset{\text{ind}}{\sim} \text{GB2}\left(e^{\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\boldsymbol{\nu_i}}\right), \quad \theta_{ij} = e^{\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\boldsymbol{\nu_i}},\ i = 1, \cdots, \ell,\ j = 1, \cdots, n_i, \\ \nu_i &\overset{\text{iid}}{\sim} N(0, \sigma^2), \\ \pi(\boldsymbol{\beta}, \sigma^2) &\propto \frac{1}{(1+\sigma^2)}. \end{aligned} \tag{3.7}$$

Combining the likelihood in (3.6) and the priors in (3.7) via Bayes' theorem, we get the joint posterior density of $\boldsymbol{\beta}, \boldsymbol{\nu}, \sigma^2|\boldsymbol{y}$ as

$$\begin{aligned} \pi(\boldsymbol{\beta}, \boldsymbol{\nu}, \sigma^2|\boldsymbol{y}) &\propto f(\boldsymbol{y}|\alpha = 3, \boldsymbol{\beta}, \boldsymbol{\nu}, \sigma^2)\, \pi(\boldsymbol{\nu}|\sigma^2)\, \pi(\boldsymbol{\beta}, \sigma^2) \\ &= \frac{\alpha^n\, e^{-\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\boldsymbol{x}'_{ij}\boldsymbol{\beta}} e^{-n_i\sum_{i=1}^{\ell}\nu_i}}{\prod_{i=1}^{\ell}\prod_{j=1}^{n_i}\left(1 + y_{ij}\, e^{-(\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\nu_i)}\right)^{\alpha+1}} \times \prod_{i=1}^{\ell}\left[\left(\frac{1}{\sigma^2}\right)^{\frac{1}{2}} e^{-\frac{\nu_i^2}{2\sigma^2}}\right] \times \frac{1}{(1+\sigma^2)^2} \\ &= \frac{1}{(1+\sigma^2)^2}\left(\frac{1}{\sigma^2}\right)^{\frac{\ell}{2}} \frac{\alpha^n\, e^{-\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\boldsymbol{x}'_{ij}\boldsymbol{\beta}} e^{-n_i\sum_{i=1}^{\ell}(n_i\nu_i+\frac{\nu_i^2}{2\sigma^2})}}{\prod_{i=1}^{\ell}\prod_{j=1}^{n_i}\left(1 + y_{ij}\, e^{-(\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\nu_i)}\right)^{\alpha+1}}. \end{aligned} \tag{3.8}$$

Let the log-likelihood function be $G(\boldsymbol{\tau}|\boldsymbol{y}) = log(f(\boldsymbol{y}|\boldsymbol{\tau}))$, where $\boldsymbol{\tau} = (\boldsymbol{\beta'}, \boldsymbol{\nu'})'$,

$$G(\boldsymbol{\tau}|\boldsymbol{y}) = n\log(\alpha) - \sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\boldsymbol{x}'_{ij}\boldsymbol{\beta} - n_i\sum_{i=1}^{\ell}\nu_i - (\alpha+1)\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\log(1 + y_{ij}\, e^{-(\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\boldsymbol{\nu_i})}).$$

For notational simplicity we write G for the log-likelihood function and then its first- and second-order partial derivatives with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\nu}$ are given by

$$\begin{aligned}
\frac{\partial G}{\partial \boldsymbol{\beta}} &= -\sum_{i=1}^{\ell}\sum_{j=1}^{n_i} \boldsymbol{x_{ij}} + (\alpha+1)\sum_{i=1}^{\ell}\sum_{j=1}^{n_i} \frac{y_{ij}\, e^{-(\boldsymbol{x'_{ij}\beta}+\nu_i)}}{1+y_{ij}\, e^{-(\boldsymbol{x'_{ij}\beta}+\nu_i)}}\, \boldsymbol{x_{ij}}, \\
\frac{\partial G}{\partial \nu_i} &= -n_i + (\alpha+1)\sum_{i=1}^{\ell}\sum_{j=1}^{n_i} \frac{y_{ij}\, e^{-(\boldsymbol{x'_{ij}\beta}+\nu_i)}}{1+y_{ij}\, e^{-(\boldsymbol{x'_{ij}\beta}+\nu_i)}}, \\
\frac{\partial^2 G}{\partial \boldsymbol{\beta}^2} &= -(\alpha+1)\sum_{i=1}^{\ell}\sum_{j=1}^{n_i} \frac{y_{ij}\, e^{-(\boldsymbol{x'_{ij}\beta}+\nu_i)}}{\left(1+y_{ij}\, e^{-(\boldsymbol{x'_{ij}\beta}+\nu_i)}\right)^2}\, \boldsymbol{x_{ij}x'_{ij}}, \\
\frac{\partial^2 G}{\partial \nu_i^2} &= -(\alpha+1)\sum_{i=1}^{\ell}\sum_{j=1}^{n_i} \frac{y_{ij}\, e^{-(\boldsymbol{x'_{ij}\beta}+\nu_i)}}{\left(1+y_{ij}\, e^{-(\boldsymbol{x'_{ij}\beta}+\nu_i)}\right)^2}, \\
\frac{\partial^2 G}{\partial \boldsymbol{\beta}\partial \nu_i} &= -(\alpha+1)\sum_{i=1}^{\ell}\sum_{j=1}^{n_i} \frac{y_{ij}\, e^{-(\boldsymbol{x'_{ij}\beta}+\nu_i)}}{\left(1+y_{ij}\, e^{-(\boldsymbol{x'_{ij}\beta}+\nu_i)}\right)^2}\, \boldsymbol{x_{ij}}.
\end{aligned}$$

Using the first-order Taylor's series approximation for $\frac{y_i\, e^{-\boldsymbol{x'_i\beta}}}{\left(1+y_i\, e^{-\boldsymbol{x'_i\beta}}\right)^2}$ at $\boldsymbol{\beta}=\mathbf{0}$, the approximate MLE of $\boldsymbol{\beta}$ is

$$\boldsymbol{\beta}^*|\alpha,\boldsymbol{\nu} = \left[\sum_{i=1}^{\ell}\sum_{j=1}^{n_i} \frac{y_{ij}\, e^{-\nu_i}}{\left(1+y_{ij}\, e^{-\nu_i}\right)^2}(\boldsymbol{x_i x'_i})\right]^{-1} \left(\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\left(\frac{y_{ij}\, e^{-\nu_i}}{1+y_{ij}\, e^{-\nu_i}} - \frac{1}{1+\alpha}\right)\boldsymbol{x_{ij}}\right). \tag{3.9}$$

Similarly, using the first-order Taylor's series approximation at $\nu_i = 0$, we have the MLE of ν_i

$$\nu^*|\alpha,\boldsymbol{\beta} = \left[\sum_{j=1}^{n_i} \frac{y_{ij}\, e^{-\boldsymbol{x'_{ij}\beta}}}{\left(1+y_{ij}\, e^{-\boldsymbol{x'_{ij}\beta}}\right)^2}\right]^{-1} \left[\sum_{j=1}^{n_i}\left(\frac{y_{ij}\, e^{-\boldsymbol{x'_{ij}\beta}}}{1+y_{ij}\, e^{-\boldsymbol{x'_{ij}\beta}}}\right) - \frac{n_i}{1+\alpha}\right]. \tag{3.10}$$

Let the gradient vectors be $\nabla G(\tau^*) = (\boldsymbol{g'_\nu},\, \boldsymbol{g'_\beta})'$, where $\boldsymbol{g_\nu} = \left(\frac{\partial G}{\partial \nu_1} \;\cdots\; \frac{\partial G}{\partial \nu_\ell}\right)'|_{\boldsymbol{\nu}=\boldsymbol{\nu}^*,\, \boldsymbol{\beta}=\boldsymbol{\beta}^*}$ and $\boldsymbol{g_\beta} = \left(\frac{\partial G}{\partial \beta_0} \;\cdots\; \frac{\partial G}{\partial \beta_p}\right)'|_{\boldsymbol{\nu}=\boldsymbol{\nu}^*,\, \boldsymbol{\beta}=\boldsymbol{\beta}^*}$ and the Hessian matrix be $H(\tau^*)$ evaluated at the approximate mode values $\boldsymbol{\beta}^*$ and $\boldsymbol{\nu}^*$. Then using the second-order Taylor's series approximation, we can write the approximated likelihood function as

$$\begin{aligned}
f(\boldsymbol{y}|\boldsymbol{\beta},\boldsymbol{\nu}) &\approx e^{\left[G(\boldsymbol{\tau^*})+\frac{1}{2}(\nabla G(\tau^*))'\,(-H(\boldsymbol{\tau^*}))^{-1}\,\nabla G(\tau^*)\right]} \\
&\qquad \times (2\pi)^{\frac{p+\ell}{2}}\, \left|(-H(\tau^*))^{-1}\right|^{\frac{1}{2}}\, N\left[\tau^* + (-H(\boldsymbol{\tau^*}))^{-1}\,\nabla G(\tau^*),\, (-H(\boldsymbol{\tau^*}))^{-1}\right].
\end{aligned}$$

Where N denotes the multivariate normal distribution for the parameter $\boldsymbol{\tau} = (\boldsymbol{\beta'},\, \boldsymbol{\nu'})'$, following the *multivariate normal approximation theorem* of Chapter 1, we can write

$$\begin{pmatrix} \boldsymbol{\nu} \\ \boldsymbol{\beta} \end{pmatrix} \sim N\left\{ \begin{pmatrix} \boldsymbol{\mu}^*_{\boldsymbol{\nu}} \\ \boldsymbol{\mu}^*_{\boldsymbol{\beta}} \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma'_{12} & \Sigma_{22} \end{pmatrix} \right\},$$

where the Hessian matrix is $H = -\begin{pmatrix} A_{11} & A_{12} \\ A'_{12} & A_{22} \end{pmatrix}$. Let

$$C(\tau^*) = e^{\left[G_\alpha(\boldsymbol{\tau}^*) + \frac{1}{2}(\nabla G_\alpha(\tau^*))'(-H_\alpha(\boldsymbol{\tau}^*))^{-1}\nabla G_\alpha(\tau^*)\right]} \left|(-H_\alpha(\tau^*))^{-1}\right|^{\frac{1}{2}}.$$

Using the same notation as in Chapter 1, equations 1.3 and 1.4 for vectors and matrices and applying the *multivariate normal approximation theorem*, we can write the approximate joint posterior density as

$$\begin{aligned}
&f(\boldsymbol{\beta}, \boldsymbol{\nu}, \sigma^2|\boldsymbol{y}) \\
&\quad \propto C(\tau^*) \times N\left(\boldsymbol{\mu}^*_\beta,\ \Sigma_{22}\right) \times N\left(\boldsymbol{\mu}^*_\nu + \Sigma_{12}\Sigma_{22}^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu}^*_\beta),\ \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma'_{12}\right) \\
&\qquad \times N\left(\mathbf{0},\ \sigma^2 I_\ell\right) \times \frac{1}{(1+\sigma^2)^2} \\
&\quad = \frac{C(\tau^*)}{(1+\sigma^2)^2} \times \frac{|A_{11}|^{\frac{1}{2}}}{|\Sigma_{22}|^{\frac{1}{2}}\ |\sigma^2 I_\ell|^{\frac{1}{2}}} \times e^{-\frac{1}{2}\left[(\boldsymbol{\beta}-\boldsymbol{\mu}^*_\beta)'\,\Sigma_{22}^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu}^*_\beta)\right]} \\
&\qquad \times e^{-\frac{1}{2}\left[(\boldsymbol{\mu}^*_\nu - A_{11}^{-1}A_{12}(\boldsymbol{\beta}-\boldsymbol{\mu}^*_\beta))'A_{11}\left(\left(A_{11}+(\sigma^2 I_\ell\right)^{-1}\right)^{-1}\left(\sigma^2 I_\ell\right)^{-1}(\boldsymbol{\mu}^*_\nu - A_{11}^{-1}A_{12}(\boldsymbol{\beta}-\boldsymbol{\mu}^*_\beta))\right]} \qquad (3.11) \\
&\qquad \times e^{-\frac{1}{2}\left[\left[\boldsymbol{\nu}-\left(A_{11}+(\sigma^2 I_\ell)^{-1}\right)^{-1}(A_{11}\boldsymbol{\mu}^*_\nu - A_{12}(\boldsymbol{\beta}-\boldsymbol{\mu}^*_\beta))\right]'\left(A_{11}+(\sigma^2 I_\ell)^{-1}\right)\left[\boldsymbol{\nu}-\left(A_{11}+(\sigma^2 I_\ell)^{-1}\right)^{-1}(A_{11}\boldsymbol{\mu}^*_\nu - A_{12}(\boldsymbol{\beta}-\boldsymbol{\mu}^*_\beta))\right]\right]}.
\end{aligned}$$

From the above joint posterior density function (3.11) we see that $\boldsymbol{\nu}$ has multivariate normal distribution given by

$$\boldsymbol{\nu}|\boldsymbol{\beta}, \sigma^2 \sim N\left[\left(A_{11} + (\sigma^2 I_\ell)^{-1}\right)^{-1}(A_{11}\boldsymbol{\mu}^*_\nu - A_{12}(\boldsymbol{\beta} - \boldsymbol{\mu}^*_\beta)),\ \left(A_{11} + (\sigma^2 I_\ell)^{-1}\right)^{-1}\right]. \quad (3.12)$$

There are numerous small areas, so we integrate out random area effects. After integrating out $\boldsymbol{\nu}$ from (3.11), we have the joint density function $\boldsymbol{\beta}, \sigma^2|\boldsymbol{y}$ as

$$\begin{aligned}
&f(\boldsymbol{\beta}, \sigma^2|\boldsymbol{y}) \\
&\quad \propto C_\alpha(\tau^*) \times \frac{1}{(1+\sigma^2)^2}\frac{|A_{11}|^{\frac{1}{2}}\ |A_{11} + (\sigma^2 I_\ell)^{-1}|^{-\frac{1}{2}}}{|\Sigma_{22}|^{\frac{1}{2}}\ |\sigma^2 I_\ell|^{\frac{1}{2}}} \times e^{-\frac{1}{2}\left[(\boldsymbol{\beta}-\boldsymbol{\mu}^*_\beta)'\,\Sigma_{22}^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu}^*_\beta)\right]} \\
&\qquad \times e^{-\frac{1}{2}\left[(\boldsymbol{\beta}-\widetilde{\boldsymbol{\mu}}_\beta)'\widetilde{\Sigma}(\boldsymbol{\beta}-\widetilde{\boldsymbol{\mu}}_\beta) - \widetilde{\boldsymbol{\mu}}'_\beta\widetilde{\Sigma}\widetilde{\boldsymbol{\mu}}_\beta + (\boldsymbol{\mu}^*_\nu + A_{11}^{-1}A_{12}\boldsymbol{\mu}^*_\beta)'S(\boldsymbol{\mu}^*_\nu + A_{11}^{-1}A_{12}\boldsymbol{\mu}^*_\beta)\right]},
\end{aligned}$$

where

$$S = A_{11}\left(A_{11} + (\sigma^2 I_\ell)^{-1}\right)^{-1}(\sigma^2 I_\ell)^{-1},$$
$$\widetilde{\boldsymbol{\mu}}_\beta = (A_{12}' A_{11}^{-1} S A_{11}^{-1} A_{12})^{-1} A_{12}' A_{11}^{-1} S \boldsymbol{\mu}_\nu^* + \boldsymbol{\mu}_\beta^*,$$
$$\widetilde{\Sigma}_\beta = A_{12}' A_{11}^{-1} S A_{11}^{-1} A_{12}.$$

From the above joint density of $\boldsymbol{\beta}, \sigma^2$ we notice that $\boldsymbol{\beta}$ has a multivariate normal distribution given by

$$\boldsymbol{\beta}|\alpha, \sigma^2, \boldsymbol{y} \sim N\left[\left(\Sigma_{22}^{-1} + \widetilde{\Sigma}_\beta\right)^{-1}\left(\Sigma_{22}^{-1}\boldsymbol{\mu}_\beta^* + \widetilde{\Sigma}_\beta \widetilde{\boldsymbol{\mu}}_\beta\right), \left(\Sigma_{22}^{-1} + \widetilde{\Sigma}_\beta\right)^{-1}\right]. \tag{3.13}$$

Integrating out $\boldsymbol{\beta}$ from the above joint density function, we get the marginal distribution of $\sigma^2|\boldsymbol{y}$

$$\begin{aligned}
&\pi(\sigma^2|\boldsymbol{y}) \\
&\quad\propto \frac{1}{(1+\sigma^2)^2} \frac{|A_{11} + (\sigma^2 I_\ell)^{-1}|^{-\frac{1}{2}} \left|\Sigma_{22}^{-1} + \widetilde{\Sigma}_\beta\right|^{-\frac{1}{2}}}{|\sigma^2 I_\ell|^{\frac{1}{2}}} \\
&\quad\quad \times e^{-\frac{1}{2}\left[(\boldsymbol{\mu}_\beta^* - \widetilde{\boldsymbol{\mu}}_\beta)' \Sigma_{22}^{-1}\left(\Sigma_{22}^{-1} + \widetilde{\Sigma}_\beta \widetilde{\boldsymbol{\mu}}_\beta\right)^{-1} \widetilde{\Sigma}_\beta (\boldsymbol{\mu}_\beta^* - \widetilde{\boldsymbol{\mu}}_\beta)\right]} \\
&\quad\quad \times e^{-\frac{1}{2}\left[-\widetilde{\boldsymbol{\mu}}_\beta' \widetilde{\Sigma}_\beta \widetilde{\boldsymbol{\mu}}_\beta + (\boldsymbol{\mu}_\nu^* + A_{11}^{-1} A_{12} \boldsymbol{\mu}_\beta^*)' S (\boldsymbol{\mu}_\nu^* + A_{11}^{-1} A_{12} \boldsymbol{\mu}_\beta^*)\right]}.
\end{aligned} \tag{3.14}$$

3.2.1 Sampling from Joint Posterior Density

We can draw approximations $\boldsymbol{\beta}$ and $\boldsymbol{\nu}$ from a multivariate normal distribution. However, the marginal distribution of $\sigma^2|\boldsymbol{y}$ is not in closed form. We use grid sampling and the Metropolis–Hastings algorithm to draw samples.

(i) Draw $\sigma^2|\boldsymbol{y}$. The conditional posterior density (3.14) is not in simple form. We used the grid sampling method. Since domain of $\sigma^2 \in (0, \infty)$, we transform σ^2 into η, which has range $(0, 1)$ and with the relation $\eta = \frac{\sigma^2}{1+\sigma^2}$. We took 1,000 grids and computed transformed probability $\pi(\eta|\boldsymbol{y})$ for each grid point. Then we draw 1,000 samples using probability distribution of grids with replacement. The samples η are transformed back to σ^2.

(ii) Using the information from the $\sigma^2|\boldsymbol{y}$ drawn above, we can draw $\boldsymbol{\beta}|\sigma^2, \boldsymbol{y}$. The Metropolis–Hastings algorithm is then used to draw jointly $\boldsymbol{\beta}, \sigma^2|\boldsymbol{y}$. The proposal distributions are t-distributions. We take the log-transformation for the proposal of σ^2. We consider $\log(\sigma^2)|\boldsymbol{y}$ to be the univariate t-distribution with d degrees of freedom, $\log(\sigma^2) \sim \mathrm{t}_d(\mu_{ln}, \sigma_{ln}^2)$, where μ_{ln} and σ_{ln}^2 are estimated from the above step. The proposal distribution for $\boldsymbol{\beta}|\boldsymbol{y}, \sigma^2$ is a multivariate t-distribution with d degrees of freedom with corresponding mean and covariance matrix as in equation (3.13). The target density is

$$
\begin{aligned}
&\pi(\boldsymbol{\beta}, \sigma^2|\boldsymbol{y}) \\
&\quad \propto \frac{\alpha^n \, e^{-\sum_{i=1}^{\ell}\sum_{j=1}^{n_i} \boldsymbol{x}'_{ij}\boldsymbol{\beta}}}{(1+\sigma^2)^2} \prod_{i=1}^{\ell} \left[\int_{\nu_i} \frac{e^{-n_i \nu_i}}{\prod_{j=1}^{n_i} \left(1 + y_{ij}\, e^{-(\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\nu_i)}\right)^{\alpha+1}} \left(\frac{1}{\sigma^2}\right)^{\frac{1}{2}} e^{-\frac{\nu_i^2}{2\sigma^2}} \, d\nu_i \right].
\end{aligned}
$$

This integration is not in simple form. We divide the integration domain into m equal intervals $[t_k,\, t_{k-1}]$ and apply a numerical integration

$$
\begin{aligned}
&\pi(\boldsymbol{\beta}, \sigma^2|\boldsymbol{y}) \\
&\qquad \propto \frac{\alpha^n \, e^{-\sum_{i=1}^{\ell}\sum_{j=1}^{n_i} \boldsymbol{x}'_{ij}\boldsymbol{\beta}}}{(1+\sigma^2)^2} \\
&\qquad\quad \times \prod_{i=1}^{\ell} \left[\sum_{k=1}^{m} \int_{t_{k-1}}^{t_k} \frac{e^{-n_i \nu_i}}{\prod_{j=1}^{n_i} \left(1 + y_{ij}\, e^{-(\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\nu_i)}\right)^{\alpha+1}} \times \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{\nu_i^2}{2\sigma^2}} \, d\nu_i \right].
\end{aligned}
$$

Using the assumption that ν_i has a univariate normal distribution centered at zero, we transform ν_i to the standard normal distribution, $z_i = \frac{\nu_i}{\sigma}$. For numerical integration we take the middle point of each interval, $\hat{z}_k = \frac{t_{k-1}+t_k}{2}$. It gives

$$\pi(\boldsymbol{\beta}, \sigma^2|\boldsymbol{y})$$

$$
\begin{aligned}
&\propto \frac{\alpha^n \, e^{-\sum_{i=1}^{\ell}\sum_{j=1}^{n_i} \boldsymbol{x}_{ij}'\boldsymbol{\beta}}}{(1+\sigma^2)^2} \\
&\quad \times \prod_{i=1}^{\ell}\left[\sum_{k=1}^{m} \frac{e^{-n_i \hat{z}_k \sigma}}{\prod_{j=1}^{n_i}\left(1+y_{ij}\, e^{-(\boldsymbol{x}_{ij}'\boldsymbol{\beta}+\hat{z}_k\sigma)}\right)^{\alpha+1}} \times \int_{t_{k-1}}^{t_k} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz\right] \\
&= \frac{\alpha^n \, e^{-\sum_{i=1}^{\ell}\sum_{j=1}^{n_i} \boldsymbol{x}_{ij}'\boldsymbol{\beta}}}{(1+\sigma^2)^2} \, e^{-\sum_{i=1}^{\ell}\sum_{j=1}^{n_i} \boldsymbol{x}_{ij}'\boldsymbol{\beta}} \\
&\quad \times \prod_{i=1}^{\ell}\left[\sum_{k=1}^{m} \frac{e^{-n_i \hat{z}_k \sigma}}{\prod_{j=1}^{n_i}\left(1+y_{ij}\, e^{-(\boldsymbol{x}_{ij}'\boldsymbol{\beta}+\hat{z}_k\sigma)}\right)^{\alpha+1}} \times (\Phi(t_k)-\Phi(t_{k-1}))\right].
\end{aligned}
$$

In the MH-algorithm MCMC sequence we keep the new sample only when it moves. We check the acceptance rate of the MH algorithm and test the convergence of the MCMC sequence.

(iii) Parameters $\nu_i|\boldsymbol{\beta}, \sigma^2$ are drawn using the Metropolis–Hastings algorithm. The proposal density is a t-distribution with d degrees of freedom. We take mean and variance for the proposal from the samples of ν_i while drawing jointly $\boldsymbol{\beta}$ and σ^2 in the above step. The target density is

$$
\pi(\nu_i|\boldsymbol{\beta}, \sigma^2) \propto \frac{e^{-\left(n_i\nu_i + \frac{\nu_i^2}{2\sigma^2}\right)}}{\prod_{j=1}^{n_i}\left(1+y_{ij}\, e^{-(\boldsymbol{x}_{ij}'\boldsymbol{\beta}+\nu_i)}\right)^{\alpha+1}}, \quad i = 1, \cdots, \ell.
$$

In the MH algorithm for drawing random area effects we check the acceptance rate of the samples. If their acceptance rate is between 0.25 and 0.50 then we keep the sample from the MH algorithm, and otherwise we discard them and in the second attempt we draw ν_i from the grid sampling method.

3.2.2 Prediction

After drawing a set of $\boldsymbol{\beta}, \boldsymbol{\nu}$, and σ^2 parameters from the GB2 model, we predict response variables as follows:

(i) Find the rate parameters θ. We calculate the rate parameter using the information on random area effect ν_i and $\boldsymbol{\beta}$ as follows

$$\theta_{ij} = e^{\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\nu_i}.$$

(ii) Find the rate parameters λ_{ij} from the gamma distribution,

$$\lambda_{ij} \sim \text{Gamma}(\alpha = 3, \theta_{ij}).$$

(iii) Predict the responses from the exponential distribution,

$$\hat{y}_{ij} \sim \text{Expo}(\lambda_{ij}).$$

3.3 Two Gamma Mixture GB2 Model Without Random Area Effects

Here we assume that the response variable has GB2 distribution as a mixture of two gamma distributions. Let both the response variable $Y|\lambda, \alpha \sim \text{Gamma}(\lambda, \alpha)$ and its rate parameter $\lambda|\phi, \theta \sim \text{Gamma}(\phi, \theta)$ have the gamma distributions

$$\begin{aligned} f(y|\alpha, \lambda) &= \frac{e^{-\lambda y}\, y^{\alpha-1}}{\Gamma(\alpha)}\, \lambda^{\alpha}, \quad \alpha, \lambda > 0, \quad \text{and} \\ f(\lambda|\phi, \theta) &= \frac{e^{-\theta\lambda}\, \lambda^{\phi-1}}{\Gamma(\phi)}\, \theta^{\phi}, \quad \alpha, \theta > 0. \end{aligned}$$

Mixing these two gamma distributions and integrating out λ, we get the GB2 density, the mixture of the two gamma distributions

$$f(y|\alpha, \phi, \theta) = \frac{y^{\alpha-1}}{\text{B}(\alpha, \phi)} \frac{1}{\theta^{\alpha}(1 + \frac{y}{\theta})^{\alpha+\phi}}, \quad \alpha, \phi, \theta > 0, \tag{3.15}$$

where $\text{B}(\alpha, \phi) = \frac{\Gamma(\alpha)\Gamma(\phi)}{\Gamma(\alpha+\phi)}$ is the beta function. The k^{th} moment of the response variable is given by

$$E[Y^k|\alpha, \phi, \theta] = \frac{\Gamma(\alpha + k)}{\Gamma\alpha} \frac{\Gamma(\phi - k)}{\Gamma(\phi)}\, \theta^k, \quad \phi > k. \tag{3.16}$$

It shows from (3.16) that we need $\phi > 2$ for variance to exist. Also, distinct α and ϕ are non-identifiable as discussed in Chapter 1, subsection *"Non-identifiable parameters in GB2* *distribution"*. To overcome non-identifiable parameters, we consider shape parameters of two levels of GB2 to be linearly related as $\phi = \alpha + 2$, and then our GB2 density function is

$$f(y|\alpha,\theta) = \frac{y^{\alpha-1}}{B(\alpha,\alpha+2)} \frac{1}{\theta^{\alpha}(1+\frac{y}{\theta})^{2(\alpha+1)}}, \quad \theta,\alpha > 0. \tag{3.17}$$

For simplicity, we write only GB2 afterward in this section to indicate the GB2 distribution as a mixture of the two gamma distribution without random area effects and linear relation of the shape parameters as defined above. We assume that the responses $y_i|\alpha,\theta,\ i = 1,\cdots,n$ are random samples from the GB2 distribution, with rate $e^{\boldsymbol{x_i'\beta}}$. The likelihood function is

$$\pi(y_i|\alpha,\boldsymbol{\beta}) = \prod_{i=1}^{n} \frac{y_i^{\alpha-1}}{B(\alpha,\alpha+2)} \frac{e^{-\alpha\,\boldsymbol{x_i'\beta}}}{\left(1+y_i\,e^{-\boldsymbol{x_i'\beta}}\right)^{2(\alpha+1)}}.$$

For Bayesian modeling, we consider a non-informative prior for $\boldsymbol{\beta}$ and α. The GB2 model without random area effects is

$$\begin{aligned} y_i|\alpha,\boldsymbol{\beta} &\overset{\text{ind}}{\sim} \text{GB2}(\alpha, e^{\boldsymbol{x_i'\beta}}), \quad \theta_i = e^{\boldsymbol{x_i'\beta}},\ i=1,\cdots,n, \\ \pi(\boldsymbol{\beta},\alpha) &\propto \frac{1}{(1+\alpha)^2}. \end{aligned} \tag{3.18}$$

The joint posterior density function is

$$\begin{aligned} \pi(\boldsymbol{\beta},\alpha|\boldsymbol{y}) &\propto f(\boldsymbol{y}|\alpha,\boldsymbol{\beta})\,\pi(\boldsymbol{\beta},\alpha) \\ &= \left[\frac{g^{\alpha-1}}{B(\alpha,\alpha+2)}\right]^{n} \frac{e^{-\sum_{i=1}^{n}\boldsymbol{x_i'\beta}}}{\prod_{i=1}^{n}\left(1+y_i\,e^{-\boldsymbol{x_i'\beta}}\right)^{2(\alpha+1)}} \times \frac{1}{(1+\alpha)^2}. \end{aligned} \tag{3.19}$$

Let the log-likelihood function be $G(\boldsymbol{\beta},\alpha|\boldsymbol{y}) = log(f(\boldsymbol{y}|\boldsymbol{\beta},\boldsymbol{\alpha}))$, and

$$G(\boldsymbol{\beta},\alpha|\boldsymbol{y}) = n\left[\alpha(\alpha-1)\log(g) - \log(\mathrm{B}(\alpha,\alpha+2))\right] - \alpha\textstyle\sum_{i=1}^{n}\boldsymbol{x_i'\beta} - 2(\alpha+1)\textstyle\sum_{i=1}^{n}\log(1+y_i\,e^{-\boldsymbol{x_i'\beta}}).$$

For notational simplicity write G for the log-likelihood function, and then its first- and second-order partial derivatives with respect to $\boldsymbol{\beta}$ are

$$\begin{aligned} \frac{\partial G}{\partial \boldsymbol{\beta}} &= \sum_{i=1}^{n}\left[-\alpha\ +\ 2(\alpha+1)\,\frac{y_i\,e^{-\boldsymbol{x_i'\beta}}}{1+y_i\,e^{-\boldsymbol{x_i'\beta}}}\right]\boldsymbol{x_i}, \\ \frac{\partial^2 G}{\partial \boldsymbol{\beta}^2} &= -\,2(\alpha+1)\sum_{i=1}^{n}\frac{y_i\,e^{-\boldsymbol{x_i'\beta}}}{\left(1+y_i\,e^{-\boldsymbol{x_i'\beta}}\right)^2}\,\boldsymbol{x_i x_i'}. \end{aligned}$$

Using the first-order Taylor's series approximation for $\frac{y_i\, e^{-\boldsymbol{x_i'\beta}}}{\left(1+y_i\, e^{-\boldsymbol{x_i'\beta}}\right)^2}$ at $\boldsymbol{\beta} = \mathbf{0}$, the approximate MLE of $\boldsymbol{\beta}$ is

$$\boldsymbol{\beta}^*|\alpha = \left[\sum_{i=1}^{n} \frac{y_i}{(1+y_i)^2} \boldsymbol{x_i x_i'}\right]^{-1} \left(\sum_{i=1}^{n} \left(\frac{y_i}{1+y_i} - \frac{\alpha}{2(\alpha+1)}\right) \boldsymbol{x_i}\right). \tag{3.20}$$

Let the gradient vectors and the Hessian matrix evaluated at the approximate mode values $\boldsymbol{\beta}^*$ be $\nabla G(\beta^*)$ and $H(\beta^*)$ respectively. Using the *multivariate normal approximation theorem* of Chapter 1 we can write the approximated joint posterior distribution as

$$\begin{aligned} f(\boldsymbol{\beta}, \alpha|\boldsymbol{y}) \propto & \frac{1}{(1+\alpha)^2}\, e^{\left[G(\boldsymbol{\beta^*}) + \frac{1}{2}(\nabla G(\boldsymbol{\beta^*}))'\,(-H(\boldsymbol{\beta^*}))^{-1}\,\nabla G(\boldsymbol{\beta^*})\right]} \\ & \times \left|(-H(\boldsymbol{\beta^*}))^{-1}\right|^{\frac{1}{2}} \mathrm{N}\left[\beta^* + (-H(\boldsymbol{\beta^*}))^{-1}\,\nabla G(\boldsymbol{\beta^*}),\, (-H(\boldsymbol{\beta^*}))^{-1}\right], \end{aligned} \tag{3.21}$$

where N is notation for the multivariate normal distribution. From the above distribution it follows that $\boldsymbol{\beta}$ has the multivariate normal distribution

$$\boldsymbol{\beta}|\alpha, \boldsymbol{y} \sim \mathrm{N}\left[\beta^* + (-H(\boldsymbol{\beta^*}))^{-1}\,\nabla G(\boldsymbol{\beta^*}),\, (-H(\boldsymbol{\beta^*}))^{-1}\right]. \tag{3.22}$$

Integrating out $\boldsymbol{\beta}$ we get marginal distribution of $\alpha|\boldsymbol{y}$ as

$$f(\alpha|\boldsymbol{y}) \propto \frac{1}{(1+\alpha)^2}\, e^{\left[G(\boldsymbol{\beta^*}) + \frac{1}{2}(\nabla G(\boldsymbol{\beta^*}))'\,(-H(\boldsymbol{\beta^*}))^{-1}\,\nabla G(\boldsymbol{\beta^*})\right]} \times \left|(-H(\boldsymbol{\beta^*}))^{-1}\right|^{\frac{1}{2}}. \tag{3.23}$$

3.3.1 Sampling from Joint Posterior Density

We can draw approximate $\boldsymbol{\beta}$ from a multivariate normal distribution. However, the marginal distribution of $\alpha|\boldsymbol{y}$ is not in a closed form. We use grid sampling and the Metropolis–Hastings algorithm to draw samples.

(i) We draw $\alpha|\boldsymbol{y}$ using the grid sampling method. Since $\alpha \in (0, \infty)$, we transform α into η which has range $(0, 1)$, with the relation $\eta = \frac{\sigma^2}{1+\sigma^2}$. We took 1,000 grid points and for each grid point computed $\boldsymbol{\beta}^*$, $G(\boldsymbol{\beta^*})$, $\nabla G(\boldsymbol{\beta^*})$ and the Hessian H to get probability distribution of $\eta|\boldsymbol{y}$ from (3.23). We draw 1,000 samples with replacement from the grid probability distribution.

(ii) Using the information from the $\alpha|\boldsymbol{y}$ sample drawn above, we can draw $\boldsymbol{\beta}|\alpha, \boldsymbol{y}$.The

Metropolis–Hastings sampling method is used to draw jointly $\boldsymbol{\beta}, \alpha|\boldsymbol{y}$. The proposal distributions are t-distributions. We take the log-transformation for the proposal of α. We consider $\log(\alpha)|\boldsymbol{y}$ as the univariate t-distribution with d degrees of freedom, $log(\alpha) \sim \mathrm{t}_d(\mu_{ln}, \sigma^2_{ln})$, where μ_{ln} and σ^2_{ln} are estimated from the above step. The proposal distribution for $\boldsymbol{\beta}|\boldsymbol{y}, \alpha$ is a multivariate t-distribution with d degrees of freedom with corresponding mean and covariance matrix given by (3.22). The target density is the joint posterior distribution function (3.19).

3.3.2 Prediction

After drawing a set of $\boldsymbol{\beta}$ and α parameters from the GB2 model, we predict response variables as follows:

(i) Find the rate parameters

$$\theta_i = e^{\boldsymbol{x'}_i \boldsymbol{\beta}}.$$

(ii) Find the rate parameters λ_i from the gamma distribution

$$\lambda_i \sim \mathrm{Gamma}(\alpha + 2, \theta_i).$$

(iii) Predict the responses from the gamma distribution

$$\hat{y}_i \sim \mathrm{Gamma}(\alpha, \lambda_i).$$

3.4 Two Gamma Mixture GB2 Model With Random Area Effects

In this section we discuss the GB2 model with random area effects as the mixture of two gamma distributions assuming noisy responses. The distribution of the noisy response variable $Y|\alpha, \lambda \sim \mathrm{Gamma}(\alpha, \lambda)$ and its rate parameter $\lambda|\phi, \theta \sim \mathrm{Gamma}(\phi, \theta)$. As discussed in Section 3.3 on response data without random area effects, we have considerd two shape parameters α and ϕ to be linearly related, $\phi = \alpha + 2$, and that variance exists in GB2 and parameters are estimable. Hereafter in this section we write only GB2 to denote the GB2 distribution as the mixture of the two gamma distributions and shape parameters related as stated above. We assume that the responses $y_{ij}|\alpha, \theta, i = 1, \cdots, \ell,\ j = 1, \cdots, n_i$ are random samples from the GB2 distribution and introduce covariates through the rate parameters as $e^{\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\boldsymbol{\nu_i}}$, where ν_i follows the normal distribution with mean zero and variance σ^2. The likelihood function is

$$\pi(\boldsymbol{y}|\alpha, \boldsymbol{\beta}, \boldsymbol{\nu}) = \prod_{i=1}^{\ell}\prod_{j=1}^{n_i} \frac{y_{ij}^{\alpha-1}}{\mathrm{B}(\alpha, \alpha+2)} \frac{e^{-\alpha(\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\nu_i)}}{\left(1 + y_{ij}\, e^{-(\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\nu_i)}\right)^{2(\alpha+1)}}. \tag{3.24}$$

Let $\boldsymbol{\beta}$ and α have non-informative independent priors. The hierarchical Bayesian GB2 model with random area effects is

$$\begin{aligned} y_{ij}|\boldsymbol{\beta}, \nu_i &\overset{\text{ind}}{\sim} \text{GB2}\left(\alpha, e^{\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\boldsymbol{\nu_i}}\right), \quad \theta_{ij} = e^{\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\boldsymbol{\nu_i}},\ i = 1, \cdots, \ell,\ j = 1, \cdots, n_i, \\ \nu_i &\overset{\text{iid}}{\sim} N(0, \sigma^2), \\ \pi(\boldsymbol{\beta}, \alpha, \sigma^2) &\propto \frac{1}{(1+\alpha)^2\,(1+\sigma^2)^2}. \end{aligned} \tag{3.25}$$

Combining the likelihood in (3.24) and the priors in (3.25) via Bayes' theorem, we get the joint posterior density of $\boldsymbol{\beta}, \alpha, \boldsymbol{\nu}, \sigma^2|\boldsymbol{y}$ as

$$\begin{aligned} \pi(\boldsymbol{\beta}, \alpha, \boldsymbol{\nu}, \sigma^2|\boldsymbol{y}) &\propto f(\boldsymbol{y}|\alpha, \boldsymbol{\beta}, \boldsymbol{\nu}, \sigma^2)\,\pi(\boldsymbol{\nu}|\sigma^2)\pi(\boldsymbol{\beta}, \alpha, \sigma^2) \\ &= \left[\frac{g^{\alpha-1}}{\mathrm{B}(\alpha, \alpha+2)}\right]^n e^{-\alpha\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\boldsymbol{x}'_{ij}\boldsymbol{\beta}} \prod_{i=1}^{\ell}\left[\frac{e^{-\alpha\, n_i\nu_i}}{\prod_{j=1}^{n_i}\left(1 + y_{ij}\, e^{-(\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\nu_i)}\right)^{2(\alpha+1)}}\right] \\ &\quad \times \prod_{i=1}^{\ell}\left[\left(\frac{1}{\sigma^2}\right)^{\frac{1}{2}} e^{-\frac{\nu_i^2}{2\sigma^2}}\right] \times \frac{1}{(1+\sigma^2)^2\,(1+\alpha)^2}. \end{aligned} \tag{3.26}$$

Let the log-likelihood function be $G(\alpha, \boldsymbol{\beta}, \boldsymbol{\nu}|\boldsymbol{y}) = log(f(\boldsymbol{y}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\nu}))$, and

$$\begin{aligned} G(\alpha, \boldsymbol{\beta}, \boldsymbol{\nu}|\boldsymbol{y}) = n\left[(\alpha-1)\log(g) - \log(\mathrm{B}(\alpha, \alpha+2))\right] - \alpha\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\boldsymbol{x}'_{ij}\boldsymbol{\beta} \\ - \sum_{i=1}^{\ell}\left[\alpha\, n_i\nu_i + 2(\alpha+1)\sum_{j=1}^{n_i}\log(1 + y_{ij}\, e^{-(\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\boldsymbol{\nu_i})})\right]. \end{aligned}$$

For notational simplicity we write G for the log-likelihood function, and then its first- and second-order partial derivatives with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\nu}$ are

$$\begin{aligned}
\frac{\partial G}{\partial \boldsymbol{\beta}} &= -\alpha \sum_{i=1}^{\ell}\sum_{j=1}^{n_i} \boldsymbol{x_{ij}} + 2(\alpha+1)\sum_{i=1}^{\ell}\sum_{j=1}^{n_i} \frac{y_{ij}\, e^{-(\boldsymbol{x'_{ij}}\boldsymbol{\beta}+\nu_i)}}{1+y_{ij}\, e^{-(\boldsymbol{x'_{ij}}\boldsymbol{\beta}+\nu_i)}}\, \boldsymbol{x_{ij}}, \\
\frac{\partial G}{\partial \nu_i} &= -\alpha\, n_i + 2(\alpha+1)\sum_{j=1}^{n_i} \frac{y_{ij}\, e^{-(\boldsymbol{x'_{ij}}\boldsymbol{\beta}+\nu_i)}}{1+y_{ij}\, e^{-(\boldsymbol{x'_{ij}}\boldsymbol{\beta}+\nu_i)}}, \\
\frac{\partial^2 G}{\partial \boldsymbol{\beta}^2} &= -2(\alpha+1)\sum_{i=1}^{\ell}\sum_{j=1}^{n_i} \frac{y_{ij}\, e^{-(\boldsymbol{x'_{ij}}\boldsymbol{\beta}+\nu_i)}}{\left(1+y_{ij}\, e^{-(\boldsymbol{x'_{ij}}\boldsymbol{\beta}+\nu_i)}\right)^2}\, \boldsymbol{x_{ij}}\boldsymbol{x'_{ij}}, \\
\frac{\partial^2 G}{\partial \nu_i^2} &= -2(\alpha+1)\sum_{j=1}^{n_i} \frac{y_{ij}\, e^{-(\boldsymbol{x'_{ij}}\boldsymbol{\beta}+\nu_i)}}{\left(1+y_{ij}\, e^{-(\boldsymbol{x'_{ij}}\boldsymbol{\beta}+\nu_i)}\right)^2}, \\
\frac{\partial^2 G}{\partial \boldsymbol{\beta}\partial \nu_i} &= -2(\alpha+1)\sum_{j=1}^{n_i} \frac{y_{ij}\, e^{-(\boldsymbol{x'_{ij}}\boldsymbol{\beta}+\nu_i)}}{\left(1+y_{ij}\, e^{-(\boldsymbol{x'_{ij}}\boldsymbol{\beta}+\nu_i)}\right)^2}\, \boldsymbol{x_{ij}}.
\end{aligned}$$

Using the first-order Taylor's series approximation for $\frac{y_i\, e^{-\boldsymbol{x'_i}\boldsymbol{\beta}}}{\left(1+y_i\, e^{-\boldsymbol{x'_i}\boldsymbol{\beta}}\right)^2}$ at $\boldsymbol{\beta} = \mathbf{0}$, the approximate MLE of $\boldsymbol{\beta}$ is

$$\boldsymbol{\beta}^*|\alpha, \boldsymbol{\nu} = \left[\sum_{i=1}^{\ell}\sum_{j=1}^{n_i} \frac{y_{ij}\, e^{-\nu_i}}{\left(1+y_{ij}\, e^{-\nu_i}\right)^2}(\boldsymbol{x_i}\boldsymbol{x'_i})\right]^{-1} \left(\sum_{i=1}^{\ell}\sum_{j=1}^{n_i} \left(\frac{y_{ij}\, e^{-\nu_i}}{1+y_{ij}\, e^{-\nu_i}} - \frac{\alpha}{2(\alpha+1)}\right)\boldsymbol{x_{ij}}\right) \tag{3.27}$$

Similarly, using the first-order Taylor's series approximation at $\nu_i = 0$, we have the MLE of ν_i given by

$$\nu^*|\alpha, \boldsymbol{\beta} = \left[\sum_{j=1}^{n_i} \frac{y_{ij}\, e^{-\boldsymbol{x'_{ij}}\boldsymbol{\beta}}}{\left(1+y_{ij}\, e^{-\boldsymbol{x'_{ij}}\boldsymbol{\beta}}\right)^2}\right]^{-1} \left[\sum_{j=1}^{n_i}\left(\frac{y_{ij}\, e^{-\boldsymbol{x'_{ij}}\boldsymbol{\beta}}}{1+y_{ij}\, e^{-\boldsymbol{x'_{ij}}\boldsymbol{\beta}}}\right) - \frac{\alpha}{2(\alpha+1)}\right] \tag{3.28}$$

Let the gradient vectors be $\nabla G(\tau^*) = (\boldsymbol{g'_\nu},\, \boldsymbol{g'_\beta})'$, where $\boldsymbol{g_\nu} = \left(\frac{\partial G}{\partial \nu_1} \ \cdots \ \frac{\partial G}{\partial \nu_\ell}\right)'|_{\boldsymbol{\nu}=\boldsymbol{\nu}^*,\, \boldsymbol{\beta}=\boldsymbol{\beta}^*}$ and $\boldsymbol{g_\beta} = \left(\frac{\partial G}{\partial \beta_0} \ \cdots \ \frac{\partial G}{\partial \beta_p}\right)'|_{\boldsymbol{\nu}=\boldsymbol{\nu}^*,\, \boldsymbol{\beta}=\boldsymbol{\beta}^*}$ and the Hessian matrix be H evaluated at the approximate mode values $\boldsymbol{\beta}^*$ and $\boldsymbol{\nu}^*$. Then using the second-order Taylor's series approximation, we can write the approximated likelihood function as

$$\begin{aligned}
f(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\nu}) &\approx e^{\left[G(\boldsymbol{\tau}^*)+\frac{1}{2}(\nabla G(\tau^*))'\,(-H(\boldsymbol{\tau}^*))^{-1}\,\nabla G(\tau^*)\right]} \\
&\quad \times (2\pi)^{\frac{p+\ell}{2}}\, \left|(-H(\tau^*))^{-1}\right|^{\frac{1}{2}}\, N\left[\tau^* + (-H(\boldsymbol{\tau}^*))^{-1}\,\nabla G(\tau^*),\, (-H(\boldsymbol{\tau}^*))^{-1}\right],
\end{aligned}$$

where N denotes the multivariate normal distribution for the parameter set $\boldsymbol{\tau} = (\boldsymbol{\beta'},\, \boldsymbol{\nu'})'$.

Following the *multivariate normal approximation theorem* of Chapter 1 we can write

$$\begin{pmatrix} \boldsymbol{\nu} \\ \boldsymbol{\beta} \end{pmatrix} \sim N\left\{ \begin{pmatrix} \boldsymbol{\mu}^*_{\boldsymbol{\nu}} \\ \boldsymbol{\mu}^*_{\boldsymbol{\beta}} \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma'_{12} & \Sigma_{22} \end{pmatrix} \right\},$$

where the Hessian matrix is $H = -\begin{pmatrix} A_{11} & A_{12} \\ A'_{12} & A_{22} \end{pmatrix}$. Let us denote

$$C_\alpha(\tau^*) = e^{\left[G_\alpha(\boldsymbol{\tau^*}) + \frac{1}{2}(\nabla G_\alpha(\tau^*))' \, (-H_\alpha(\boldsymbol{\tau^*}))^{-1} \, \nabla G_\alpha(\tau^*)\right]} \left|(-H_\alpha(\tau^*))^{-1}\right|^{\frac{1}{2}}.$$

Using the same notation as in Chapter 1, equations 1.3 and 1.4 for vectors and matrices and applying the *multivariate normal approximation theorem* we can write the approximate joint posterior density as

$$\begin{aligned}
&f(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\nu}, \sigma^2 | \boldsymbol{y}) \\
&\quad \propto C_\alpha(\tau^*) \times N\left(\boldsymbol{\mu}^*_\beta, \ \Sigma_{22}\right) \times N\left(\boldsymbol{\mu}^*_\nu + \Sigma_{12}\Sigma_{22}^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu}^*_\beta), \ \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma'_{12}\right) \\
&\qquad \times N\left(\mathbf{0}, \ \sigma^2 I_\ell\right) \times \frac{1}{(1+\sigma^2)^2 \, (1+\alpha)^2} \\
&\quad = C_\alpha(\tau^*) \times \frac{1}{(1+\sigma^2)^2 \, (1+\alpha)^2} \\
&\qquad \times \frac{|A_{11}|^{\frac{1}{2}}}{|\Sigma_{22}|^{\frac{1}{2}} \ |\sigma^2 I_\ell|^{\frac{1}{2}}} \times e^{-\frac{1}{2}\left[(\boldsymbol{\beta} - \boldsymbol{\mu}^*_\beta)' \, \Sigma_{22}^{-1} \, (\boldsymbol{\beta} - \boldsymbol{\mu}^*_\beta)\right]} \\
&\qquad \times e^{-\frac{1}{2}\left[(\boldsymbol{\mu}^*_\nu - A_{11}^{-1}A_{12}(\boldsymbol{\beta} - \boldsymbol{\mu}^*_\beta))' A_{11}\left(\left(A_{11} + (\sigma^2 I_\ell\right)^{-1}\right)^{-1}\left(\sigma^2 I_\ell\right)^{-1}(\boldsymbol{\mu}^*_\nu - A_{11}^{-1}A_{12}(\boldsymbol{\beta} - \boldsymbol{\mu}^*_\beta))\right]} \\
&\qquad \times e^{-\frac{1}{2}\left[\left[\boldsymbol{\nu} - \left(A_{11} + (\sigma^2 I_\ell)^{-1}\right)^{-1}(A_{11}\boldsymbol{\mu}^*_\nu - A_{12}(\boldsymbol{\beta} - \boldsymbol{\mu}^*_\beta))\right]'\left(A_{11} + (\sigma^2 I_\ell)^{-1}\right)\left[\boldsymbol{\nu} - \left(A_{11} + (\sigma^2 I_\ell)^{-1}\right)^{-1}(A_{11}\boldsymbol{\mu}^*_\nu - A_{12}(\boldsymbol{\beta} - \boldsymbol{\mu}^*_\beta))\right]\right]}.
\end{aligned} \tag{3.29}$$

From the above joint posterior density function we see that $\boldsymbol{\nu}$ has multivariate normal distribution

$$\boldsymbol{\nu} | \boldsymbol{\beta}, \alpha, \sigma^2 \sim N\left[\left(A_{11} + (\sigma^2 I_\ell)^{-1}\right)^{-1}\left(A_{11}\boldsymbol{\mu}^*_\nu - A_{12}(\boldsymbol{\beta} - \boldsymbol{\mu}^*_\beta)\right), \left(A_{11} + (\sigma^2 I_\ell)^{-1}\right)^{-1}\right]. \tag{3.30}$$

There are numerous small areas. Integrating out $\boldsymbol{\nu}$, we have the joint density function of $\boldsymbol{\beta}, \alpha, \sigma^2 | \boldsymbol{y}$ given by

$$f(\boldsymbol{\beta}, \alpha, \sigma^2 | \boldsymbol{y})$$

$$\propto C_\alpha(\tau^*) \times \frac{|A_{11}+(\sigma^2 I_\ell)^{-1}|^{-\frac{1}{2}}}{(1+\sigma^2)^2\,(1+\alpha)^2} \times \frac{|A_{11}|^{\frac{1}{2}}}{|\Sigma_{22}|^{\frac{1}{2}}\ |\sigma^2 I_\ell|^{\frac{1}{2}}} \times e^{-\frac{1}{2}\left[(\boldsymbol{\beta}-\boldsymbol{\mu}_\beta^*)'\,\Sigma_{22}^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu}_\beta^*)\right]}$$
$$\times\, e^{-\frac{1}{2}\left[(\boldsymbol{\beta}-\widetilde{\boldsymbol{\mu}}_\beta)'\widetilde{\Sigma}(\boldsymbol{\beta}-\widetilde{\boldsymbol{\mu}}_\beta)-\widetilde{\boldsymbol{\mu}}_\beta'\widetilde{\Sigma}\widetilde{\boldsymbol{\mu}}_\beta+(\boldsymbol{\mu}_\nu^*+A_{11}^{-1}A_{12}\boldsymbol{\mu}_\beta^*)'S(\boldsymbol{\mu}_\nu^*+A_{11}^{-1}A_{12}\boldsymbol{\mu}_\beta^*)\right]},$$

where

$$S = A_{11}\left(A_{11}+(\sigma^2 I_\ell)^{-1}\right)^{-1}(\sigma^2 I_\ell)^{-1},$$
$$\widetilde{\boldsymbol{\mu}}_\beta = (A_{12}'A_{11}^{-1}SA_{11}^{-1}A_{12})^{-1}A_{12}'A_{11}^{-1}S\boldsymbol{\mu}_\nu^* + \boldsymbol{\mu}_\beta^*,$$
$$\widetilde{\Sigma}_\beta = A_{12}'A_{11}^{-1}SA_{11}^{-1}A_{12}.$$

From the above joint density of $\boldsymbol{\beta}, \alpha, \sigma^2|\boldsymbol{y}$, we notice that $\boldsymbol{\beta}$ has a multivariate normal distribution

$$\boldsymbol{\beta}|\alpha,\sigma^2,\boldsymbol{y} \sim N\left[\left(\Sigma_{22}^{-1}+\widetilde{\Sigma}_\beta\right)^{-1}\left(\Sigma_{22}^{-1}\boldsymbol{\mu}_\beta^*+\widetilde{\Sigma}_\beta\widetilde{\boldsymbol{\mu}}_\beta\right),\ \left(\Sigma_{22}^{-1}+\widetilde{\Sigma}_\beta\right)^{-1}\right]. \tag{3.31}$$

Integrating out $\boldsymbol{\beta}$ from the above joint density function, we get the joint distribution of $\alpha, \sigma^2|\boldsymbol{y}$

$$\pi(\alpha,\sigma^2|\boldsymbol{y})$$
$$\propto C_\alpha(\tau^*)\frac{1}{(1+\sigma^2)^2\,(1+\alpha)^2}\frac{|A_{11}+(\sigma^2 I_\ell)^{-1}|^{-\frac{1}{2}}\left|\Sigma_{22}^{-1}+\widetilde{\Sigma}_\beta\right|^{-\frac{1}{2}}}{|\sigma^2 I_\ell|^{-\frac{1}{2}}}$$
$$\times\, e^{-\frac{1}{2}\left[(\boldsymbol{\mu}_\beta^*-\widetilde{\boldsymbol{\mu}}_\beta)'\Sigma_{22}^{-1}\left(\Sigma_{22}^{-1}+\widetilde{\Sigma}_\beta\widetilde{\boldsymbol{\mu}}_\beta\right)^{-1}\widetilde{\Sigma}_\beta(\boldsymbol{\mu}_\beta^*-\widetilde{\boldsymbol{\mu}}_\beta)\right]}$$
$$\times\, e^{-\frac{1}{2}\left[-\widetilde{\boldsymbol{\mu}}_\beta'\widetilde{\Sigma}_\beta\widetilde{\boldsymbol{\mu}}_\beta+(\boldsymbol{\mu}_\nu^*+A_{11}^{-1}A_{12}\boldsymbol{\mu}_\beta^*)'S(\boldsymbol{\mu}_\nu^*+A_{11}^{-1}A_{12}\boldsymbol{\mu}_\beta^*)\right]}. \tag{3.32}$$

3.4.1 Sampling from Joint Posterior Density

The joint posterior density function of α and σ^2 is not in the simple form. We borrow parameter α from the previous hierarchical Bayesian GB2 model, *two gamma mixture GB2 model without random area effects* and use it as an approximation. We can draw $\boldsymbol{\beta}$ from a multivariate normal distribution. However, the marginal density function of $\sigma^2|\boldsymbol{y}$ is not in simple form. We use the grid sampling method and MH algorithm to draw parameters.

(i) We borrow α from the previous section 3.3, *two gamma mixture GB2 model without* *random area effects*. From 1,000 samples of α from that model we pick 100 quantile values as an approximation from the previous model without random area effects.

(ii) We draw $\sigma^2|\alpha, \boldsymbol{y}$ using the grid sampling method with marginal density function

$$\begin{aligned}
&\pi(\sigma^2|\alpha,\boldsymbol{y}) \\
&\qquad \propto \frac{1}{(1+\sigma^2)^2} \frac{|A_{11} + (\sigma^2 I_\ell)^{-1}|^{-\frac{1}{2}} \left|\Sigma_{22}^{-1} + \widetilde{\Sigma}_\beta\right|^{-\frac{1}{2}}}{|\sigma^2 I_\ell|^{\frac{1}{2}}} \\
&\qquad \times e^{-\frac{1}{2}\left[(\boldsymbol{\mu}_\beta^* - \widetilde{\boldsymbol{\mu}}_\beta)' \Sigma_{22}^{-1} \left(\Sigma_{22}^{-1} + \widetilde{\Sigma}_\beta\right)^{-1} \widetilde{\Sigma}_\beta (\boldsymbol{\mu}_\beta^* - \widetilde{\boldsymbol{\mu}}_\beta)\right]} \\
&\qquad \times e^{-\frac{1}{2}\left[-\widetilde{\boldsymbol{\mu}}_\beta' \widetilde{\Sigma}_\beta \widetilde{\boldsymbol{\mu}}_\beta + (\boldsymbol{\mu}_\nu^* + A_{11}^{-1} A_{12} \boldsymbol{\mu}_\beta^*)' S (\boldsymbol{\mu}_\nu^* + A_{11}^{-1} A_{12} \boldsymbol{\mu}_\beta^*)\right]}. \qquad (3.33)
\end{aligned}$$

The domain of $\sigma^2 \in (0, \infty)$. So we transform σ^2 into η which has range $(0, 1)$, with realtion $\eta = \frac{\sigma^2}{1+\sigma^2}$. We took 100 grid values of η and computed the transformed probability $\pi(\eta|\alpha, \boldsymbol{y})$ from (3.33). For each quantile value of η we draw one α and transform it back to σ^2.

(iii) Using the information α and $\sigma^2|\alpha$ drawn above, we can draw $\boldsymbol{\beta}|\alpha, \sigma^2, \boldsymbol{y}$. The Metropolis–Hastings algorithm is used to draw jointly $\boldsymbol{\beta}, \alpha, \sigma^2|\boldsymbol{y}$. The proposal distributions are t-distributions.

We take the log-transformation for the proposal of α and σ^2. We consider $\log(\alpha, \sigma^2)|\boldsymbol{y}$ is the bivariate t-distribution with d degrees of freedom, $\log(\alpha, \sigma^2) \sim \mathrm{t}_d(\boldsymbol{\mu}_{ln}, \sigma_{ln}^2)$, where $\boldsymbol{\mu}_{ln}$ and Σ_{ln} are estimated from the above step. The proposal distribution for $\boldsymbol{\beta}|\alpha, \sigma^2, \boldsymbol{y}$ is a multivariate t-distribution with d degrees of freedom with corresponding mean and covariance matrix as in equation (3.31). The target density is

$$\begin{aligned}
&\pi(\boldsymbol{\beta}, \alpha, \sigma^2|\boldsymbol{y}) \\
&\qquad \propto \frac{e^{-\alpha \sum_{i=1}^{\ell} \sum_{j=1}^{n_i} \boldsymbol{x}_{ij}' \boldsymbol{\beta}}}{(1+\sigma^2)^2 (1+\alpha)^2} \left[\frac{g^{\alpha-1}}{\mathrm{B}(\alpha, \alpha+2)}\right]^n \\
&\qquad \times \prod_{i=1}^{\ell} \left[\int_{\nu_i} \frac{e^{-\alpha n_i \nu_i}}{\prod_{j=1}^{n_i} \left(1 + y_{ij}\, e^{-(\boldsymbol{x}_{ij}'\boldsymbol{\beta} + \nu_i)}\right)^{2(\alpha+1)}} \left(\frac{1}{\sigma^2}\right)^{\frac{1}{2}} e^{-\frac{\nu_i^2}{2\sigma^2}}\, d\nu_i\right].
\end{aligned}$$

This integration is not in simple form. We divide the integration domain into m equal intervals $[t_k,\, t_{k-1}]$ and apply a numerical integration

$$\pi(\boldsymbol{\beta}, \alpha, \sigma^2|\boldsymbol{y}) \propto \frac{e^{-\alpha\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\boldsymbol{x}'_{ij}\boldsymbol{\beta}}}{(1+\sigma^2)^2\,(1+\alpha)^2} \left[\frac{g^{\alpha-1}}{\mathrm{B}(\alpha,\alpha+2)}\right]^n \times \prod_{i=1}^{\ell}\left[\sum_{k=1}^{m}\int_{t_{k-1}}^{t_k}\frac{e^{-\alpha\, n_i\nu_i}}{\prod_{j=1}^{n_i}\left(1+y_{ij}\,e^{-(\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\nu_i)}\right)^{2(\alpha+1)}}\times\frac{1}{\sqrt{2\pi\sigma^2}}\,e^{-\frac{\nu_i^2}{2\sigma^2}}\,d\nu_i\right].$$

Using the assumption that ν_i has a univariate normal distribution centered at zero, we transform ν_i to the standard normal distribution, $z_i = \frac{\nu_i}{\sigma}$. For numerical integration we take the middle point of each interval, $\hat{z}_k = \frac{t_{k-1}+t_k}{2}$. It gives

$$\begin{aligned}\pi(\boldsymbol{\beta}, \alpha, \sigma^2|\boldsymbol{y}) &\propto \frac{e^{-\alpha\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\boldsymbol{x}'_{ij}\boldsymbol{\beta}}}{(1+\sigma^2)^2\,(1+\alpha)^2} \left[\frac{g^{\alpha-1}}{\mathrm{B}(\alpha,\alpha+2)}\right]^n \\ &\quad\times \prod_{i=1}^{\ell}\left[\sum_{k=1}^{m}\frac{e^{-\alpha\, n_i\hat{z}_k\sigma}}{\prod_{j=1}^{n_i}\left(1+y_{ij}\,e^{-(\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\hat{z}_k\sigma)}\right)^{2(\alpha+1)}}\times\int_{t_{k-1}}^{t_k}\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}dz\right] \\ &= \frac{e^{-\alpha\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\boldsymbol{x}'_{ij}\boldsymbol{\beta}}}{(1+\sigma^2)^2\,(1+\alpha)^2} \left[\frac{g^{\alpha-1}}{\mathrm{B}(\alpha,\alpha+2)}\right]^n \\ &\quad\times \prod_{i=1}^{\ell}\left[\sum_{k=1}^{m}\frac{e^{-n_i\hat{z}_k\sigma}}{\prod_{j=1}^{n_i}\left(1+y_{ij}\,e^{-(\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\hat{z}_k\sigma)}\right)^{\alpha+1}}\times(\Phi(t_k)-\Phi(t_{k-1}))\right].\end{aligned}$$

In the MH-algorithm MCMC sequence we keep the new sample only when it moves. We check the acceptance rate of the MH algorithm and test the convergence of the MCMC sequence.

(iv) Parameters $\nu_i|\boldsymbol{\beta}, \alpha, \sigma^2$ are drawn using the Metropolis–Hastings algorithm. The proposal density is a t-distribution with d degrees of freedom. We take the mean and variance for the proposal from the samples of ν_i while drawing $\boldsymbol{\beta}, \alpha$, and σ^2 jointly in the above step. The target density is

$$\pi(\nu_i|\boldsymbol{\beta}, \alpha, \sigma^2) \propto \frac{e^{-\left(\alpha\, n_i\nu_i+\frac{\nu_i^2}{2\sigma^2}\right)}}{\prod_{j=1}^{n_i}\left(1+y_{ij}\,e^{-(\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\nu_i)}\right)^{2(\alpha+1)}}, \quad i = 1, \cdots, \ell.$$

We keep the samples from the MH algorithm if their acceptance rate is between 0.25

and 0.50, and otherwise we discard them. If their acceptance rate is not between 0.25 and 0.50 then in the second attempt we draw ν_i using the grid sampling method.

3.4.2 Prediction

After drawing a set of $\alpha, \boldsymbol{\beta}, \boldsymbol{\nu}$, and σ^2 parameters from the GB2 model, we predict response variables as follows:

(i) Find the rate parameters θ. We calculate the rate parameter using the information on random area effect ν_i and $\boldsymbol{\beta}$ as follows

$$\theta_{ij} = e^{\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\nu_i}.$$

(ii) Find the rate parameters λ_{ij} from the gamma distribution

$$\lambda_{ij} \sim \text{Gamma}(\alpha = \alpha + 2, \theta_{ij}).$$

(iii) Predict the responses from the gamma distribution

$$\hat{y}_{ij} \sim \text{Gamma}(\alpha, \lambda_{ij}).$$

3.5 Two Generalized Gamma Mixture GB2 Model Without Random Area Effects

In this section we assume that the response variable has a mixture of two generalized gamma distributions without random effects. Let the response variable $Y|\alpha, \lambda, \gamma \sim \text{GGamma}(\alpha, \lambda, \gamma)$, and its rate parameter $\lambda|\phi, \theta, \gamma \sim \text{GGamma}(\phi, \theta, \gamma)$ both having the generalized gamma distribution

$$f(y|\alpha, \lambda, \gamma) = \gamma \frac{e^{-(\lambda y)^\gamma}\, y^{\alpha-1}}{\Gamma(\frac{\alpha}{\gamma})} \lambda^\alpha, \quad \alpha, \lambda, \gamma > 0,$$
$$f(\lambda|\phi, \theta, \gamma) = \gamma \frac{e^{-(\theta\lambda)^\gamma}\, \lambda^{\phi-1}}{\Gamma(\frac{\phi}{\gamma})} \theta^\phi, \quad \phi, \theta, \gamma > 0.$$

Mixing these two distributions and integrating out rate parameter λ, we get the mixture of two generalized gamma GB2 distributions

$$f(y|\alpha,\phi,\theta,\gamma) = \frac{\gamma}{\mathrm{B}\left(\frac{\alpha}{\gamma},\frac{\phi}{\gamma}\right)} \frac{y^{\alpha-1}\,\theta^{-\alpha}}{\left(1+(\frac{y}{\theta})^{\gamma}\right)^{\frac{\alpha+\phi}{\gamma}}}, \quad \alpha,\phi,\theta,\gamma>0,$$

where $\mathrm{B}\left(\frac{\alpha}{\gamma}.\frac{\phi}{\gamma}\right) = \frac{\Gamma(\frac{\alpha}{\gamma})\,\Gamma(\frac{\phi}{\gamma})}{\Gamma\left(\frac{\alpha+\phi}{\gamma}\right)}$ is the beta function. The k^{th} moment of the response is given by

$$E[Y^k|\alpha,\phi,\theta] = \frac{\Gamma\left(\frac{\alpha+k}{\gamma}\right)}{\Gamma\left(\frac{\alpha}{\gamma}\right)} \frac{\Gamma\left(\frac{\phi-k}{\gamma}\right)}{\Gamma\left(\frac{\phi}{\gamma}\right)}\,\theta^k, \quad \phi>k. \tag{3.34}$$

It shows from (3.34) that we need $\phi>2$ for the variance to exist. Also, distinct α and ϕ are non-identifiable as discussed in Chapter 1, subsection *"Non-identifiable parameters in GB2 distribution"*. To overcome a non-identifiable problem, we consider the shape parameters of two levels of GB2 to be linearly related as $\phi=\alpha+2$, and then our GB2 density function is

$$f(y|\alpha,\theta,\gamma) = \frac{\gamma}{\mathrm{B}\left(\frac{\alpha}{\gamma},\frac{\alpha+2}{\gamma}\right)} \frac{y^{\alpha-1}\,\theta^{-\alpha}}{\left(1+(\frac{y}{\theta})^{\gamma}\right)^{\frac{2(\alpha+1)}{\gamma}}}, \quad \alpha,\theta,\gamma>0.$$

For simplicity, we write only GB2 afterward in this section to indicate the GB2 distribution as a mixture of the two generalized gamma distribution without random area effects and relation of the shape parameters as defined above. We assume that the responses $y_i|\alpha,\theta,\gamma,\ i=1,\cdots,n$ are random samples from the GB2 distribution, the mixture of the two generalized gammas, and covariates are introduced through the rate parameter as $e^{\boldsymbol{x}_i'\boldsymbol{\beta}}$. The likelihood function is

$$\pi(\boldsymbol{y}|\alpha,\boldsymbol{\beta},\gamma) = \prod_{i=1}^{n} \frac{\gamma}{\mathrm{B}\left(\frac{\alpha}{\gamma},\frac{\alpha+2}{\gamma}\right)} \frac{y_i^{\alpha-1}\,e^{-\alpha\,\boldsymbol{x}_i'\boldsymbol{\beta}}}{\left[1+(y_i\,e^{-\boldsymbol{x}_i'\boldsymbol{\beta}})^{\gamma}\right]^{\frac{2(\alpha+1)}{\gamma}}}, \quad \alpha,\theta,\gamma>0.$$

For Bayesian modeling, we consider a non-informative prior for $\boldsymbol{\beta}$ and α and an informative prior for γ. The GB2 model without random area effects is

$$\begin{aligned} y_i|\alpha, \boldsymbol{\beta}, \gamma &\overset{\text{ind}}{\sim} \text{GB2}(\alpha, e^{\boldsymbol{x_i'\beta}}, \gamma), \quad \theta_i = e^{\boldsymbol{x_i'\beta}},\ i = 1, \cdots, n, \\ \pi(\boldsymbol{\beta}, \alpha) &\propto \frac{1}{(1+\alpha)^2}, \\ \gamma &\sim \text{Gamma}(S, R), \quad \text{where shape S and rate R are specified.} \end{aligned} \tag{3.35}$$

The joint posterior density function is

$$\pi(\alpha, \boldsymbol{\beta}, \gamma|\boldsymbol{y}) = \frac{e^{-R\gamma}\,\gamma^{S-1}}{(1+\alpha)^2}\left[\frac{\gamma g^{\alpha-1}}{\text{B}\left(\frac{\alpha}{\gamma}, \frac{\alpha+2}{\gamma}\right)}\right]^n \frac{e^{-\alpha\sum_{i=1}^n \boldsymbol{x_i'\beta}}}{\prod_{i=1}^n \left[1 + (y_i\, e^{-\boldsymbol{x_i'\beta}})^\gamma\right]^{\frac{2(\alpha+1)}{\gamma}}}. \tag{3.36}$$

Let the log-likelihood function be $G(\alpha, \boldsymbol{\beta}, \gamma|\boldsymbol{y}) = log(f(\boldsymbol{y|\alpha, \beta, \gamma}))$, and

$$\begin{aligned} G(\alpha, \boldsymbol{\beta}, \gamma|\boldsymbol{y}) = n&\left[\log(\gamma) + (\alpha-1)\log(g) - \log\left(\text{B}\left(\frac{\alpha}{\gamma}, \frac{\alpha+2}{\gamma}\right)\right)\right] \\ &- \alpha\sum_{i=1}^n \boldsymbol{x_i'\beta} - \frac{2(\alpha+1)}{\gamma}\sum_{i=1}^n \log\left(1 + (y_i\, e^{-\boldsymbol{x_i'\beta}})^\gamma\right). \end{aligned}$$

For notational simplicity we write G for the log-likelihood functionl, and then its first- and second-order partial derivatives with respect to $\boldsymbol{\beta}$ are

$$\begin{aligned} \frac{\partial G}{\partial \boldsymbol{\beta}} &= \sum_{i=1}^n \left[-\alpha\ +\ 2(\alpha+1)\,\frac{(y_i\, e^{-\boldsymbol{x_i'\beta}})^\gamma}{1 + (y_i\, e^{-\boldsymbol{x_i'\beta}})^\gamma}\right]\boldsymbol{x_i}, \\ \frac{\partial^2 G}{\partial \boldsymbol{\beta}^2} &= -\ 2(\alpha+1)\,\gamma\sum_{i=1}^n \frac{(y_i\, e^{-\boldsymbol{x_i'\beta}})^\gamma}{\left(1 + (y_i\, e^{-\boldsymbol{x_i'\beta}})^\gamma\right)^2}\,\boldsymbol{x_i x_i'}. \end{aligned}$$

Using the first-order Taylor's series approximation for $\frac{(y_i\, e^{-\boldsymbol{x_i'\beta}})^\gamma}{1+(y_i\, e^{-\boldsymbol{x_i'\beta}})^\gamma}$ at $\boldsymbol{\beta} = \boldsymbol{0}$, the approximate MLE of $\boldsymbol{\beta}$ is

$$\boldsymbol{\beta}^*|\alpha = \left[\gamma\sum_{i=1}^n \frac{y_i^\gamma}{(1+y_i^\gamma)^2}\boldsymbol{x_i x_i'}\right]^{-1}\left(\sum_{i=1}^n\left(\frac{y_i^\gamma}{1+y_i^\gamma} - \frac{\alpha}{2(\alpha+1)}\right)\boldsymbol{x_i}\right). \tag{3.37}$$

Let the gradient vectors and the Hessian matrix evaluated at the approximate mode values $\boldsymbol{\beta}^*$ be $\nabla G(\beta^*)$ and $H(\beta^*)$ respectively. Then using the second-order Taylor's series approximation, we can write the approximated likelihood function as

$$\begin{aligned} f(\boldsymbol{\beta}, \alpha, \gamma|\boldsymbol{y}) \propto\ & \frac{e^{-R\gamma}\,\gamma^{S-1}}{(1+\sigma^2)^2}\, e^{\left[G(\boldsymbol{\beta^*}) + \frac{1}{2}(\nabla G(\boldsymbol{\beta^*}))'\,(-H(\boldsymbol{\beta^*}))^{-1}\,\nabla G(\boldsymbol{\beta^*})\right]} \\ & \times \left|(-H(\boldsymbol{\beta^*}))^{-1}\right|^{\frac{1}{2}}\,\text{N}\left[\beta^* + (-H(\boldsymbol{\beta^*}))^{-1}\,\nabla G(\boldsymbol{\beta^*}),\ (-H(\boldsymbol{\beta^*}))^{-1}\right]. \end{aligned} \tag{3.38}$$

From the above distribution it follows that $\boldsymbol{\beta}$ has the multivariate normal distribution given by

$$\boldsymbol{\beta}|\alpha, \gamma, \boldsymbol{y} \sim \mathrm{N}\left[\beta^* + (-H(\boldsymbol{\beta}^*))^{-1}\, \nabla G(\boldsymbol{\beta}^*),\ (-H(\boldsymbol{\beta}^*))^{-1}\right]. \tag{3.39}$$

Integrating out $\boldsymbol{\beta}$ we get the joint distribution of $\alpha, \gamma|\boldsymbol{y}$ given by

$$f(\alpha, \gamma|\boldsymbol{y}) \propto \tfrac{e^{-R\gamma}\,\gamma^{S-1}}{(1+\sigma^2)^2}\, e^{\left[G(\boldsymbol{\beta}^*)+\frac{1}{2}(\nabla G(\boldsymbol{\beta}^*))'\,(-H(\boldsymbol{\beta}^*))^{-1}\,\nabla G(\boldsymbol{\beta}^*)\right]} \times |(-H(\boldsymbol{\beta}^*))^{-1}|^{\frac{1}{2}} \tag{3.40}$$

3.5.1 Sampling from Joint Posterior Density

We pick shape and rate parameter $S = R = 1$ for our prior distributioin of $\gamma \sim \text{Gamma}(S, R)$. We can draw an approximate $\boldsymbol{\beta}$ from a multivariae normal distribution. However, the joint posterior density of α and γ is not in simple form. We used grid sampling and the Metropolis–Hastings algorithm to draw parameters.

(i) We borrow α from the previous model, *two gamma mixture GB2 model with random area effects*. We pick 100 quantile values from 1,000 samples from that model and used them as approximated samples.

(ii) We draw γ using the grid sampling method for given α from the above step. Since $\gamma \in (0, \infty)$, we transform $\eta = \frac{\gamma}{1+\gamma}$ so that $\eta \in (0, 1)$. We make 100 grids for η and calculate transformed density $\eta|\alpha, \boldsymbol{y}$ from (3.40). For each given α we draw a sample using grid probability distribution of $\eta|\alpha$ and then transform it back to γ.

(iii) Using the information from $\alpha, \gamma|\boldsymbol{y}$ samples drawn above, we can draw $\boldsymbol{\beta}|\alpha, \gamma, \boldsymbol{y}$. The Metropolis–Hastings algorithm is then used to draw $\boldsymbol{\beta}, \alpha, \gamma|\boldsymbol{y}$ jointly. The proposal distributions are t-distributions. We take the log-transformation for the proposal of α and γ. We consider $\log(\alpha, \gamma)|\boldsymbol{y}$ is the bivariate t-distribution with d degrees of freedom, $\log(\alpha, \gamma) \sim \mathrm{t}_d(\boldsymbol{\mu}_{ln}, \Sigma_{ln})$, where μ_{ln} and Σ_{ln} are estimated from the above step. The proposal distribution for $\boldsymbol{\beta}|\boldsymbol{y}, \alpha, \gamma$ is a multivariate t-distribution with d degrees of freedom with corresponding mean and covariance matrix as in (3.39). The target density is the joint posterior distribution (3.36).

3.5.2 Prediction

We note that if $y|\alpha, \lambda, \gamma \sim \text{GGamma}(\alpha, \lambda, \gamma)$ then with transformation $t = (\lambda y)^\gamma$, we get $(\lambda y)^\gamma \sim \text{Gamma}\left(\frac{\alpha}{\gamma}, 1\right)$. We use this information for prediction of responses. After drawing a set of $\boldsymbol{\beta}, \alpha$, and γ parameters from the GB2 model, we predict responses as follows:

(i) Find the rate parameters

$$\theta_i = e^{\boldsymbol{x}_i'\boldsymbol{\beta}}.$$

(ii) Find the rate parameters λ_i. Draw a gamma random sample G_1 from the gamma distribution $\text{Gamma}\left(\frac{\alpha+2}{\gamma}, 1\right)$ then calculate λ_i

$$\lambda_i = \frac{\mathrm{G}_1^{\frac{1}{\gamma}}}{\theta_i}.$$

(iii) Predict the responses. Draw a gamma random sample G_2 from the gamma distribution $\text{Gamma}\left(\frac{\alpha}{\gamma}, 1\right)$ then predict y_i

$$\hat{y}_i = \frac{\mathrm{G}_2^{\frac{1}{\gamma}}}{\lambda_i}.$$

3.6 Two Generalized Gamma Mixture GB2 Model With Random Area Effects

In this section we discuss the GB2 model as the mixture of two generalized-gamma distributions with random area effects. The response variable $Y|\alpha, \lambda, \gamma \sim \text{GGamma}(\alpha, \lambda, \gamma)$ and its rate parameter $\lambda|\phi, \theta, \gamma \sim \text{GGamma}(\phi, \theta, \gamma)$ both have generalized gamma distribution. As discussed, the model without random area effects in section 3.5, we considered two shape parameters α and ϕ to be linearly related, $\phi = \alpha+2$, the variance of GB2 distribution to exist, and the parameters to be estimable. Here afterward in this section we write only GB2 to denote this distribution. We assume that the responses $y_{ij}|\alpha, \theta, \gamma,\ i = 1, \cdots, n,\ j = 1, \cdots, n_i$ are random samples from the GB2 distribution. We introduce the covariates through the rate parameter by writing $e^{\boldsymbol{x}_{ij}'\boldsymbol{\beta}+\nu_i}$ and assume ν_i follows the normal distribution with mean zero and variance σ^2. The likelihood function is

$$\pi(\boldsymbol{y}|\alpha,\boldsymbol{\beta},\boldsymbol{\nu}) = \prod_{i=1}^{\ell}\prod_{j=1}^{n_i}\gamma\frac{y_{ij}^{\alpha-1}}{\mathrm{B}(\frac{\alpha}{\gamma},\frac{\alpha+2}{\gamma})}\frac{e^{-\alpha(\boldsymbol{x'_{ij}\beta}+\nu_i)}}{\left(1+\left[y_{ij}\,e^{-(\boldsymbol{x'_{ij}\beta}+\nu_i)}\right]^{\gamma}\right)^{\frac{2(\alpha+1)}{\gamma}}}. \tag{3.41}$$

Let $\boldsymbol{\beta}$ and α have non-informative priors and γ have an informative prior. The priors are independent. The hierarchical Bayesian GB2 model with random area effects is

$$\begin{aligned}
y_{ij}|\boldsymbol{\beta},\alpha,\gamma,\nu_i &\overset{\text{ind}}{\sim} \mathrm{GB2}\left(\alpha, e^{\boldsymbol{x'_{ij}\beta}+\boldsymbol{\nu_i}},\gamma\right),\ \theta_{ij}=e^{\boldsymbol{x'_{ij}\beta}+\boldsymbol{\nu_i}},\ i=1,\cdots,n,\ j=1,\cdots,n_i,\\
\nu_i &\overset{\text{iid}}{\sim} N(0,\sigma^2),\\
\pi(\boldsymbol{\beta},\alpha,\sigma^2) &\propto \frac{1}{(1+\alpha)^2\,(1+\sigma^2)^2}\\
\gamma &\sim \mathrm{Gamma}(S,R), \qquad \text{where shape S and rate R are specified.}
\end{aligned} \tag{3.42}$$

Combining the likelihood in (3.41) and the priors in (4.4) via Bayes theorem, we get the joint posterior density of $\alpha,\boldsymbol{\beta},\gamma,\boldsymbol{\nu},\sigma^2|\boldsymbol{y}$ as

$$\begin{aligned}
\pi(\alpha,\boldsymbol{\beta},\gamma,\boldsymbol{\nu},\sigma^2|\boldsymbol{y}) &\propto f(\boldsymbol{y}|\alpha,\boldsymbol{\beta},\gamma,\boldsymbol{\nu},\sigma^2)\,\pi(\boldsymbol{\nu}|\sigma^2)\pi(\boldsymbol{\beta},\alpha,\sigma^2)\\
&= \left[\frac{\gamma\,g^{\alpha-1}}{\mathrm{B}(\frac{\alpha}{\gamma},\frac{\alpha+2}{\gamma})}\right]^n e^{-\alpha\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\boldsymbol{x'_{ij}\beta}}\prod_{i=1}^{\ell}\left[\frac{e^{-\alpha\,n_i\nu_i}}{\prod_{j=1}^{n_i}\left(1+\left[y_{ij}\,e^{-(\boldsymbol{x'_{ij}\beta}+\nu_i)}\right]^{\gamma}\right)^{\frac{2(\alpha+1)}{\gamma}}}\right]\\
&\times\prod_{i=1}^{\ell}\left[\left(\frac{1}{\sigma^2}\right)^{\frac{1}{2}}e^{-\frac{\nu_i^2}{2\sigma^2}}\right]\times\frac{e^{-R\gamma}\,\gamma^{S-1}}{(1+\alpha)^2\,(1+\sigma^2)^2}
\end{aligned} \tag{3.43}$$

Let the log-likelihood function be $G(\alpha,\boldsymbol{\beta},\gamma,\boldsymbol{\nu}|\boldsymbol{y}) = log(f(\boldsymbol{y}|\alpha,\boldsymbol{\beta},\gamma\boldsymbol{\nu}))$, and

$$\begin{aligned}
&G(\alpha,\boldsymbol{\beta},\gamma,\boldsymbol{\nu}|\boldsymbol{y})\\
&\quad= n\left[\log(\gamma)+(\alpha-1)\log(g)-\log\left(\mathrm{B}(\frac{\alpha}{\gamma},\frac{\alpha+2}{\gamma})\right)\right]-\alpha\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\boldsymbol{x'_{ij}\beta}\\
&\quad\quad\times-\sum_{i=1}^{\ell}\left[\alpha\,n_i\nu_i+\frac{2(\alpha+1)}{\gamma}\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\log\left(1+\left[y_{ij}\,e^{-(\boldsymbol{x'_{ij}\beta}+\nu_i)}\right]^{\gamma}\right)\right].
\end{aligned}$$

For notational simplicity we write G for the log-likelihood function and its first- and second-order partial derivatives with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\nu}$ are

$$\frac{\partial G}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\left(-\alpha\ + 2(\alpha+1)\frac{\left[y_{ij}\,e^{-(\boldsymbol{x'_{ij}}\boldsymbol{\beta}+\nu_i)}\right]^{\gamma}}{1+\left[y_{ij}\,e^{-(\boldsymbol{x'_{ij}}\boldsymbol{\beta}+\nu_i)}\right]^{\gamma}}\right)\boldsymbol{x_{ij}},$$

$$\frac{\partial G}{\partial \nu_i} = -\,\alpha\,n_i + 2(\alpha+1)\sum_{j=1}^{n_i}\frac{\left[y_{ij}\,e^{-(\boldsymbol{x'_{ij}}\boldsymbol{\beta}+\nu_i)}\right]^{\gamma}}{1+\left[y_{ij}\,e^{-(\boldsymbol{x'_{ij}}\boldsymbol{\beta}+\nu_i)}\right]^{\gamma}},$$

$$\frac{\partial^2 G}{\partial \boldsymbol{\beta}^2} = -\,2(\alpha+1)\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\gamma\frac{\left[y_{ij}\,e^{-(\boldsymbol{x'_{ij}}\boldsymbol{\beta}+\nu_i)}\right]^{\gamma}}{\left(1+\left[y_{ij}\,e^{-(\boldsymbol{x'_{ij}}\boldsymbol{\beta}+\nu_i)}\right]^{\gamma}\right)^2}\boldsymbol{x_{ij}}\boldsymbol{x'_{ij}},$$

$$\frac{\partial^2 G}{\partial \nu_i^2} = -\,2(\alpha+1)\sum_{j=1}^{n_i}\gamma\frac{\left[y_{ij}\,e^{-(\boldsymbol{x'_{ij}}\boldsymbol{\beta}+\nu_i)}\right]^{\gamma}}{\left(1+\left[y_{ij}\,e^{-(\boldsymbol{x'_{ij}}\boldsymbol{\beta}+\nu_i)}\right]^{\gamma}\right)^2},$$

$$\frac{\partial^2 G}{\partial \boldsymbol{\beta}\partial \nu_i} = -\,2(\alpha+1)\sum_{j=1}^{n_i}\gamma\frac{\left[y_{ij}\,e^{-(\boldsymbol{x'_{ij}}\boldsymbol{\beta}+\nu_i)}\right]^{\gamma}}{\left(1+\left[y_{ij}\,e^{-(\boldsymbol{x'_{ij}}\boldsymbol{\beta}+\nu_i)}\right]^{\gamma}\right)^2}\boldsymbol{x_{ij}}.$$

Using the first-order Taylor's series approximation for $\frac{\left(y_i\,e^{-(\boldsymbol{x'_i}\boldsymbol{\beta}+\boldsymbol{\nu_i})}\right)^{\gamma}}{1+\left(y_i\,e^{-(\boldsymbol{x'_i}\boldsymbol{\beta}+\boldsymbol{\nu_i})}\right)^{\gamma}}$ at $\boldsymbol{\beta}=\mathbf{0}$, the approximate MLE of $\boldsymbol{\beta}$ is

$$\boldsymbol{\beta}^*|\alpha,\gamma,\boldsymbol{\nu} = \left[\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\frac{\gamma\left(y_{ij}\,e^{-\nu_i}\right)^{\gamma}}{\left(1+\left(y_{ij}\,e^{-\nu_i}\right)^{\gamma}\right)^2}(\boldsymbol{x_i}\boldsymbol{x'_i})\right]^{-1}\left(\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\left(\frac{\left(y_{ij}\,e^{-\nu_i}\right)^{\gamma}}{1+\left(y_{ij}\,e^{-\nu_i}\right)^{\gamma}}-\frac{\alpha}{2(\alpha+1)}\right)\boldsymbol{x_{ij}}\right) \quad (3.44)$$

Similarly, using the first-order Taylor's series approximation at $\nu_i = 0$, we have the MLE of ν_i given by

$$\nu_i^*|\alpha,\gamma,\boldsymbol{\beta} = \left[\sum_{j=1}^{n_i}\frac{\gamma\left(y_{ij}\,e^{-\boldsymbol{x'_{ij}}\boldsymbol{\beta}}\right)^{\gamma}}{\left(1+\left(y_{ij}\,e^{-\boldsymbol{x'_{ij}}\boldsymbol{\beta}}\right)^{\gamma}\right)^2}\right]^{-1}\left[\sum_{j=1}^{n_i}\left(\frac{\left(y_{ij}\,e^{-\boldsymbol{x'_{ij}}\boldsymbol{\beta}}\right)^{\gamma}}{1+\left(y_{ij}\,e^{-\boldsymbol{x'_{ij}}\boldsymbol{\beta}}\right)^{\gamma}}-\frac{\alpha\,n_i}{2(\alpha+1)}\right)\right]. \quad (3.45)$$

Let the gradient vectors be $\nabla G(\tau^*) = (\boldsymbol{g'_\nu},\,\boldsymbol{g'_\beta})'$, where $\boldsymbol{g_\nu} = \left(\begin{matrix}\frac{\partial G}{\partial \nu_1} & \cdots & \frac{\partial G}{\partial \nu_\ell}\end{matrix}\right)'|_{\boldsymbol{\nu}=\boldsymbol{\nu}^*,\,\boldsymbol{\beta}=\boldsymbol{\beta}^*}$ and $\boldsymbol{g_\beta} = \left(\begin{matrix}\frac{\partial G}{\partial \beta_0} & \cdots & \frac{\partial G}{\partial \beta_p}\end{matrix}\right)'|_{\boldsymbol{\nu}=\boldsymbol{\nu}^*,\,\boldsymbol{\beta}=\boldsymbol{\beta}^*}$ and the Hessian matrix be H evaluated at the approximate mode values $\boldsymbol{\beta}^*$ and $\boldsymbol{\nu}^*$. Then using the second-order Taylor's series approximation, we can write the approximated likelihood function as

$$f(\boldsymbol{y}|\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{\nu}) \approx e^{\left[G(\boldsymbol{\tau}^*)+\frac{1}{2}(\nabla G(\tau^*))'\,(-H(\boldsymbol{\tau}^*))^{-1}\,\nabla G(\tau^*)\right]}$$
$$\times\,(2\pi)^{\frac{p+\ell}{2}}\,\left|(-H(\tau^*))^{-1}\right|^{\frac{1}{2}}\,N\left[\tau^*+(-H(\boldsymbol{\tau}^*))^{-1}\,\nabla G(\tau^*),\,(-H(\boldsymbol{\tau}^*))^{-1}\right].$$

Where N denotes the multivariate normal distribution for the parameter set $\boldsymbol{\tau} = (\boldsymbol{\beta'},\, \boldsymbol{\nu'})'$. Following the *multivariate normal approximation theorem* of Chapter 1 we can write

$$\begin{pmatrix} \boldsymbol{\nu} \\ \boldsymbol{\beta} \end{pmatrix} \sim N \left\{ \begin{pmatrix} \boldsymbol{\mu}_{\boldsymbol{\nu}}^{*} \\ \boldsymbol{\mu}_{\boldsymbol{\beta}}^{*} \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{pmatrix} \right\},$$

where the Hessian matrix is $H = -\begin{pmatrix} A_{11} & A_{12} \\ A_{12}' & A_{22} \end{pmatrix}$. Let us denote

$$C_{\alpha\gamma}(\tau^{*}) = e^{\left[G(\boldsymbol{\tau}^{*}) + \frac{1}{2}(\nabla G(\tau^{*}))'\,(-H(\boldsymbol{\tau}^{*}))^{-1}\,\nabla G(\tau^{*})\right]} \left|(-H(\tau^{*}))^{-1}\right|^{\frac{1}{2}}.$$

Using the same notation as in Chapter 1, equations 1.3 and 1.4 for vectors and matrices and applying the *multivariate normal approximation theorem* we can write the approximate joint posterior density as

$$\begin{aligned}
&f(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\nu}, \sigma^{2}|\boldsymbol{y}) \\
&\quad \propto C_{\alpha\gamma}(\tau^{*}) \times N\left(\boldsymbol{\mu}_{\beta}^{*},\; \Sigma_{22}\right) \times N\left(\boldsymbol{\mu}_{\nu}^{*} + \Sigma_{12}\Sigma_{22}^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu}_{\beta}^{*}),\; \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}'\right) \\
&\qquad \times N\left(\mathbf{0},\, \sigma^{2} I_{\ell}\right) \times \frac{e^{-R\gamma}\,\gamma^{S-1}}{(1+\alpha)^{2}\,(1+\sigma^{2})^{2}} \\
&\quad = C_{\alpha\gamma}(\tau^{*}) \times \frac{e^{-R\gamma}\,\gamma^{S-1}}{(1+\alpha)^{2}\,(1+\sigma^{2})^{2}} \times \frac{|A_{11}|^{\frac{1}{2}}}{|\Sigma_{22}|^{\frac{1}{2}}\;|\sigma^{2} I_{\ell}|^{\frac{1}{2}}} \times e^{-\frac{1}{2}\left[(\boldsymbol{\beta}-\boldsymbol{\mu}_{\beta}^{*})'\,\Sigma_{22}^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu}_{\beta}^{*})\right]} \\
&\qquad \times e^{-\frac{1}{2}\left[(\boldsymbol{\mu}_{\nu}^{*}-A_{11}^{-1}A_{12}(\boldsymbol{\beta}-\boldsymbol{\mu}_{\beta}^{*}))' A_{11}\left(\left(A_{11}+(\sigma^{2}I_{\ell}\right)^{-1}\right)^{-1}\left(\sigma^{2}I_{\ell}\right)^{-1}(\boldsymbol{\mu}_{\nu}^{*}-A_{11}^{-1}A_{12}(\boldsymbol{\beta}-\boldsymbol{\mu}_{\beta}^{*}))\right]} \\
&\qquad \times e^{-\frac{1}{2}\left[\left[\boldsymbol{\nu}-\left(A_{11}+(\sigma^{2}I_{\ell})^{-1}\right)^{-1}(A_{11}\boldsymbol{\mu}_{\nu}^{*}-A_{12}(\boldsymbol{\beta}-\boldsymbol{\mu}_{\beta}^{*}))\right]'\left(A_{11}+(\sigma^{2}I_{\ell})^{-1}\right)\left[\boldsymbol{\nu}-\left(A_{11}+(\sigma^{2}I_{\ell})^{-1}\right)^{-1}(A_{11}\boldsymbol{\mu}_{\nu}^{*}-A_{12}(\boldsymbol{\beta}-\boldsymbol{\mu}_{\beta}^{*}))\right]\right]}.
\end{aligned} \tag{3.46}$$

From the above joint posterior density function (3.46), we see that $\boldsymbol{\nu}$ has multivariate normal distribution

$$\boldsymbol{\nu}|\boldsymbol{\beta}, \alpha, \sigma^{2} \sim N\left[\left(A_{11} + (\sigma^{2}I_{\ell})^{-1}\right)^{-1}\left(A_{11}\boldsymbol{\mu}_{\nu}^{*} - A_{12}(\boldsymbol{\beta} - \boldsymbol{\mu}_{\beta}^{*})\right),\, \left(A_{11} + (\sigma^{2}I_{\ell})^{-1}\right)^{-1}\right]. \tag{3.47}$$

There are numerous small areas. Integrating out $\boldsymbol{\nu}$, we have the joint density function of $\boldsymbol{\beta}, \alpha, \gamma, \sigma^{2}|\boldsymbol{y}$ as follows

$$f(\boldsymbol{\beta}, \alpha, \gamma, \sigma^{2}|\boldsymbol{y})$$

$$\propto C_{\alpha\gamma}(\tau^*) \times \frac{e^{-R\gamma}\,\gamma^{S-1}}{(1+\alpha)^2\,(1+\sigma^2)^2} \times \frac{|A_{11}|^{\frac{1}{2}}\ |A_{11}+(\sigma^2 I_\ell)^{-1}|^{-\frac{1}{2}}}{|\Sigma_{22}|^{\frac{1}{2}}\ |\sigma^2 I_\ell|^{\frac{1}{2}}} \times e^{-\frac{1}{2}\left[(\boldsymbol{\beta}-\boldsymbol{\mu}^*_\beta)'\,\Sigma_{22}^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu}^*_\beta)\right]}$$
$$\times\, e^{-\frac{1}{2}\left[(\boldsymbol{\beta}-\widetilde{\boldsymbol{\mu}}_\beta)'\widetilde{\Sigma}(\boldsymbol{\beta}-\widetilde{\boldsymbol{\mu}}_\beta) - \widetilde{\boldsymbol{\mu}}'_\beta\widetilde{\Sigma}\widetilde{\boldsymbol{\mu}}_\beta + (\boldsymbol{\mu}^*_\nu+A_{11}^{-1}A_{12}\boldsymbol{\mu}^*_\beta)'S(\boldsymbol{\mu}^*_\nu+A_{11}^{-1}A_{12}\boldsymbol{\mu}^*_\beta)\right]},$$

where

$$S = A_{11}\left(A_{11}+(\sigma^2 I_\ell)^{-1}\right)^{-1}(\sigma^2 I_\ell)^{-1},$$
$$\widetilde{\boldsymbol{\mu}}_\beta = (A'_{12}A_{11}^{-1}SA_{11}^{-1}A_{12})^{-1}A'_{12}A_{11}^{-1}S\boldsymbol{\mu}^*_\nu + \boldsymbol{\mu}^*_\beta,$$
$$\widetilde{\Sigma}_\beta = A'_{12}A_{11}^{-1}SA_{11}^{-1}A_{12}.$$

From the above joint density of $\boldsymbol{\beta},\alpha,\gamma,\sigma^2|\boldsymbol{y}$ we notice that $\boldsymbol{\beta}$ has a multivariate normal distribution

$$\boldsymbol{\beta}|\alpha,\gamma,\sigma^2,\boldsymbol{y} \sim N\left[\left(\Sigma_{22}^{-1}+\widetilde{\Sigma}_\beta\right)^{-1}\left(\Sigma_{22}^{-1}\boldsymbol{\mu}^*_\beta+\widetilde{\Sigma}_\beta\widetilde{\boldsymbol{\mu}}_\beta\right),\ \left(\Sigma_{22}^{-1}+\widetilde{\Sigma}_\beta\right)^{-1}.\right] \qquad (3.48)$$

Integrating out $\boldsymbol{\beta}$ from the above joint density function, we get the joint density of $\alpha,\gamma,\sigma^2|\boldsymbol{y}$

$$\pi(\alpha,\gamma,\sigma^2|\boldsymbol{y})$$
$$\propto C_{\alpha\gamma}(\tau^*)\frac{e^{-R\gamma}\,\gamma^{S-1}}{(1+\alpha)^2\,(1+\sigma^2)^2}\,\frac{|A_{11}|^{\frac{1}{2}}\ |A_{11}+(\sigma^2 I_\ell)^{-1}|^{-\frac{1}{2}}\left|\Sigma_{22}^{-1}+\widetilde{\Sigma}_\beta\right|^{-\frac{1}{2}}}{|\Sigma_{22}|^{\frac{1}{2}}\ |\sigma^2 I_\ell|^{\frac{1}{2}}}$$
$$\times\, e^{-\frac{1}{2}\left[(\boldsymbol{\mu}^*_\beta-\widetilde{\boldsymbol{\mu}}_\beta)'\Sigma_{22}^{-1}\left(\Sigma_{22}^{-1}+\widetilde{\Sigma}_\beta\widetilde{\boldsymbol{\mu}}_\beta\right)^{-1}\widetilde{\Sigma}_\beta(\boldsymbol{\mu}^*_\beta-\widetilde{\boldsymbol{\mu}}_\beta)\right]} \qquad (3.49)$$
$$\times\, e^{-\frac{1}{2}\left[-\widetilde{\boldsymbol{\mu}}'_\beta\widetilde{\Sigma}_\beta\widetilde{\boldsymbol{\mu}}_\beta+(\boldsymbol{\mu}^*_\nu+A_{11}^{-1}A_{12}\boldsymbol{\mu}^*_\beta)'S(\boldsymbol{\mu}^*_\nu+A_{11}^{-1}A_{12}\boldsymbol{\mu}^*_\beta)\right]}.$$

3.6.1 Sampling from Joint Posterior Density

We pick the shape and rate parameter $S = R = 1$ for our prior distribution $\gamma \sim \text{Gamma}(S, R)$. Grid sampling and the Metropolis–Hastings sampling algorithm are used for drawing samples. For the Metropolis–Hastings algorithm we use t-distribution with d degrees of freedom as our proposal distribution.

(i) We borrow α and γ from the previous model, *two generalized gamma mixture GB2 model without random area effects*. From these samples we pick a set of 100 quantiles. For quantiles we keep the variable α and then then γ in ascending order.

(ii) We draw $\sigma^2|\alpha,\gamma,\boldsymbol{y}$ using the grid sampling method with density function given by

$$\pi(\sigma^2|\alpha,\gamma,\boldsymbol{y}) \propto \frac{1}{(1+\sigma^2)^2} \frac{|A_{11}|^{\frac{1}{2}}\ |A_{11}+(\sigma^2 I_\ell)^{-1}|^{-\frac{1}{2}} \left|\Sigma_{22}^{-1}+\widetilde{\Sigma}_\beta\right|^{-\frac{1}{2}}}{|\Sigma_{22}|^{\frac{1}{2}}\ |\sigma^2 I_\ell|^{\frac{1}{2}}}$$
$$\times\, e^{-\frac{1}{2}\left[(\boldsymbol{\mu}_\beta^* - \widetilde{\boldsymbol{\mu}}_\beta)'\Sigma_{22}^{-1}\left(\Sigma_{22}^{-1}+\widetilde{\Sigma}_\beta\right)^{-1}\widetilde{\Sigma}_\beta(\boldsymbol{\mu}_\beta^* - \widetilde{\boldsymbol{\mu}}_\beta)\right]}$$
$$\times\, e^{-\frac{1}{2}\left[-\widetilde{\boldsymbol{\mu}}_\beta'\widetilde{\Sigma}_\beta\widetilde{\boldsymbol{\mu}}_\beta + (\boldsymbol{\mu}_\nu^* + A_{11}^{-1}A_{12}\boldsymbol{\mu}_\beta^*)'S(\boldsymbol{\mu}_\nu^* + A_{11}^{-1}A_{12}\boldsymbol{\mu}_\beta^*)\right]}. \quad (3.50)$$

The domain of $\sigma^2 \in (0,\infty)$. So we transform σ^2 into η which has range $(0,1)$, $\sigma^2 = \frac{\eta}{1-\eta}$. We took 100 grid values of η and computed transformed probability $\pi(\eta|\alpha,\boldsymbol{y})$ from (3.50). For each set of quantile values of α and γ we draw η and then transform it back to σ^2.

(iii) Using the information α,γ and $\sigma^2|\alpha,\gamma$ drawn above, we can draw $\boldsymbol{\beta}|\alpha,\gamma,\sigma^2,\boldsymbol{y}$. The Metropolis–Hastings algorithm is then used to draw $\boldsymbol{\beta},\alpha,\gamma,\sigma^2|\boldsymbol{y}$ jointly. The proposal distributions are t-distributions. The proposal density for $\log(\alpha,\gamma,\sigma^2)|\boldsymbol{y}$ is the multivariabe t-distribution with d degrees of freedom, $\log(\alpha,\gamma,\sigma^2) \sim \mathrm{t}_d(\boldsymbol{\mu}_{ln},\Sigma_{ln})$, where $\boldsymbol{\mu}_{ln}$ and Σ_{ln} are estimated from the above step. The proposal distribution for $\boldsymbol{\beta}|\alpha,\gamma,\sigma^2,\boldsymbol{y}$ is a multivariate t-distribution with d degrees of freedom with corresponding mean and covariance matrix as in equation (3.48). The target density is as follows

$$\pi(\boldsymbol{\beta},\alpha,\gamma,\sigma^2|\boldsymbol{y})$$
$$\propto \frac{\gamma^{S-1}}{(1+\alpha)^2\,(1+\sigma^2)^2}\left[\frac{\gamma\, g^{\alpha-1}}{\mathrm{B}(\frac{\alpha}{\gamma},\frac{\alpha+2}{\gamma})}\right]^n e^{-(\alpha\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\boldsymbol{x}_{ij}'\boldsymbol{\beta}+R\gamma)}$$
$$\times \prod_{i=1}^{\ell}\left[\int_{\nu_i}\frac{e^{-\alpha\, n_i\nu_i}}{\prod_{j=1}^{n_i}\left(1+\left[y_{ij}\, e^{-(\boldsymbol{x}_{ij}'\boldsymbol{\beta}+\nu_i)}\right]^{\gamma}\right)^{\frac{2(\alpha+1)}{\gamma}}}\left(\frac{1}{\sigma^2}\right)^{\frac{1}{2}} e^{-\frac{\nu_i^2}{2\sigma^2}}\,d\nu_i\right].$$

This integration is not in simple form. We apply a numerical integration. We divide integration domain into m equal intervals $[t_k,\, t_{k-1}]$

$$\pi(\boldsymbol{\beta},\alpha,\gamma,\sigma^2|\boldsymbol{y})$$

$$\propto \frac{\gamma^{S-1}}{(1+\alpha)^2\,(1+\sigma^2)^2}\left[\frac{\gamma\, g^{\alpha-1}}{\mathrm{B}(\frac{\alpha}{\gamma},\frac{\alpha+2}{\gamma})}\right]^n e^{-(\alpha\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\boldsymbol{x}'_{ij}\boldsymbol{\beta}+R\gamma)}$$

$$\times\prod_{i=1}^{\ell}\left[\sum_{k=1}^{m}\int_{t_{k-1}}^{t_k}\frac{e^{-\alpha\, n_i\nu_i}}{\prod_{j=1}^{n_i}\left(1+\left[y_{ij}\, e^{-(\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\nu_i)}\right]^{\gamma}\right)^{\frac{2(\alpha+1)}{\gamma}}}\times\frac{1}{\sqrt{2\pi\sigma^2}}\,e^{-\frac{\nu_i^2}{2\sigma^2}}\,d\nu_i\right].$$

Using the assumption that ν_i has a univariate normal distribution centered at zero, we transform ν_i to the standard normal distribution, $z_i=\frac{\nu_i}{\sigma}$. For numerical integration we take the middle point of each interval, $\hat{z}_k=\frac{t_{k-1}+t_k}{2}$

$$\pi(\boldsymbol{\beta},\alpha,\gamma,\sigma^2|\boldsymbol{y})$$

$$\propto \frac{\gamma^{S-1}}{(1+\alpha)^2\,(1+\sigma^2)^2}\left[\frac{\gamma\, g^{\alpha-1}}{\mathrm{B}(\frac{\alpha}{\gamma},\frac{\alpha+2}{\gamma})}\right]^n e^{-(\alpha\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\boldsymbol{x}'_{ij}\boldsymbol{\beta}+R\gamma)}$$

$$\times\prod_{i=1}^{\ell}\left[\sum_{k=1}^{m}\frac{e^{-\alpha\, n_i\hat{z}_k\sigma}}{\prod_{j=1}^{n_i}\left(1+\left[y_{ij}\, e^{-(\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\hat{z}_k\sigma)}\right]^{\gamma}\right)^{\frac{2(\alpha+1)}{\gamma}}}\times\int_{t_{k-1}}^{t_k}\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}dz\right]$$

$$= \frac{\gamma^{S-1}}{(1+\alpha)^2\,(1+\sigma^2)^2}\left[\frac{\gamma\, g^{\alpha-1}}{\mathrm{B}(\frac{\alpha}{\gamma},\frac{\alpha+2}{\gamma})}\right]^n e^{-(\alpha\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\boldsymbol{x}'_{ij}\boldsymbol{\beta}+R\gamma)}$$

$$\times\prod_{i=1}^{\ell}\left[\sum_{k=1}^{m}\frac{e^{-n_i\hat{z}_k\sigma}}{\prod_{j=1}^{n_i}\left(1+\left[y_{ij}\, e^{-(\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\hat{z}_k\sigma)}\right]^{\gamma}\right)^{\frac{2(\alpha+1)}{\gamma}}}\times(\Phi(t_k)-\Phi(t_{k-1}))\right].$$

(iv) Parameters $\nu_i|\boldsymbol{\beta},\alpha,\gamma,\sigma^2$ are drawn using the Metropolis–Hastings algorithm. The proposal density is a t-distribution with d degrees of freedom. We take the mean and variance for the proposal from the samples of ν_i while drawing $(\boldsymbol{\beta},\alpha,\gamma,\sigma^2)$ in the above step. The target density is

$$\pi(\nu_i|\boldsymbol{\beta},\alpha,\gamma,\sigma^2)\propto\frac{e^{-\left(\alpha\, n_i\nu_i+\frac{\nu_i^2}{2\sigma^2}\right)}}{\prod_{j=1}^{n_i}\left(1+\left[y_{ij}\, e^{-(\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\nu_i)}\right]^{\gamma}\right)^{\frac{2(\alpha+1)}{\gamma}}},\quad i=1,\cdots,\ell.$$

We keep the samples ν_i from the Metropolis–Hastings algorithm if the acceptance rate lies between 0.25 and 0.50, and otherwise we discard them and sample again using the grid sampling method in the second attempt.

3.6.2 Prediction

After drawing all parameters from the GB2 distribution model as mention above, we predict the responses as follows:

(i) Find the rate parameters θ. We calculate the rate parameter using the information on random area effect ν_i and $\boldsymbol{\beta}$ as follows

$$\theta_{ij} = e^{\boldsymbol{x'}_{ij}\boldsymbol{\beta}+\nu_i}.$$

(ii) Draw shape parameters λ. In the GB2 distribution let us consider a transformation $t = (\theta\lambda)^\gamma$. This gives $(\theta\lambda)^\gamma \sim \text{Gamma}\left(\frac{\alpha+2}{\gamma}, 1\right)$. Say we draw a random sample G_1 from this distribution, then we can calculate λ as follows

$$\begin{aligned}
\mathrm{G}_1 &= (\theta\lambda)^\gamma \sim \text{Gamma}\left(\frac{\alpha+2}{\gamma}, 1\right) \\
\lambda_{ij} &= \frac{\mathrm{G}_1^{\frac{1}{\gamma}}}{\theta_{ij}}.
\end{aligned}$$

(iii) Predict responses. In the GB2 distribution we consider a transformation $t = (\lambda y)^\gamma$. This gives $(\lambda y)^\gamma \sim \text{Gamma}\left(\frac{\alpha}{\gamma}, 1\right)$. Say we draw a random sample G_2 from this distribution; then we can predict $\hat{y}$ as follows:

$$\begin{aligned}
\mathrm{G}_2 &= (\lambda_{ij} y)^\gamma \sim \text{Gamma}\left(\frac{\alpha}{\gamma}, 1\right) \\
\hat{y}_{ij} &= \frac{\mathrm{G}_2^{\frac{1}{\gamma}}}{\lambda_{ij}}.
\end{aligned}$$

3.7 Application and Results

We have applied GB2 models assuming noisy responses with nine covariates from NLSS-II. We have given the results for three special cases for GB2 models: the mixture of the exponential and gamma distributions, the mixture of two gamma distributions, and the mixture of two generalized gamma distributions.

We have sampled parameters using the grid sampling method and the MCMC Metropolis–Hastings sampling method. We have applied the MH algorithm more than once whenever needed. We have presented the acceptance rates for the final MH algorithm. We note that if we have more than one MH algorithm, the samples obtained from previous MH are also used by the final MH algorithm. For all the models fitted for noisy responses, we have taken a set of 2100 samples, "burn-in" 100 samples and thinning interval of one. The final set has 1000 samples.

For model-comparison purposes, we have calculated LPML values. The larger the value of LPML the better the model. We have also presented the percentage of CPO values below 0.02 probability for each model.

Table 3.1 presents LPML values for GB2 models with and without random area effects for all six models developed in this chapter. We have seen a trend for noiseless modeling with and without random area effects in Chapter 2; this table also shows a similar trend, that the mixture of the exponential and gamma models have much smaller LPML values compared to the mixture of two gamma and the mixture of two generalized gamma GB2 models. Therefore obviously the mixture of two gammas or the mixture of two generalized gammas GB2 model fits better in NLSS-II consumption data. For example in stratum two, the LPML values for GB2 models without random area effects: the mixture of exponential and gamma, two gamma and two generalized gamma distributions models are respectively -728.7, -623.8, and -620.5. Similarly, LPML values for the GB2 models with random area effects: the mixture of exponential and gamma, two gamma, and two generalized gamma distributions models are respectively -741.5, -614.7, and -591.8.

Comparing the GB2 models without random area effects, the LPML values for the mixture of two gamma and two generalized gamma are similar even though the mixture of the two generalized gamma has slightly higher LPML values. For example in stratum one, the LPML values for the mixture of two gamma and the mixture of two generalized gamma are respectively -221.8 and -220.1. In a similar trend, comparing the models with random area effects, the mixture of the two generalized gamma distributions has slightly higher LPML values than the mixture of two gamma distributions. For example in stratum one, the LPML values for the mixture of two gamma and the mixture of two generalized gamma are respectively -195.2 and -173.9.

In Table 3.1 we have also provided a column for LPML values for hierarchical Bayesian nested error regression models. Since this hierarchical Bayesian NER model is built with a logarithmic transformation, for possible close comparison purposes, we exponentiate back the predicted responses and calculate the probability using the log-normal distribution. This table shows that hierarchical Bayesian NER has smaller LPML than the mixture of two generalized gamma distribution GB2 model.

Table 3.2 presents the percentage of observations with CPO values below 0.02 probability for all models fitted in this chapter: the mixture of exponential and gamma, the mixture of two gamma, and the mixture of two generalized gamma GB2 models, with and without random area effects. The smaller percentage value is better. Here we see that every model without random area effects shows a percentage below 5%. From this table it is not clear whether some one is better than another model or not. However, it is still good to check if our model has a proportion of observations below some small threshold probability.

Table 3.3 presents the parameters-assignment results for the mixture of exponential and gamma GB2 model. It provides the MH sampler acceptance rate for parameters, Geweke convergence diagnostic test, and effective sample sizes. The acceptance rates of parameter $\boldsymbol{\beta}$ for the model without random area effects and parameters $(\boldsymbol{\beta}, \sigma^2)$ for the model with random area effects are provided for all strata. All Geweke test p-values are larger than 0.05 except for one, for the model without random area effects models for parameter *beta8* with 0.02. The effective sample sizes for all parameters are almost one except very few parameters have less, about 0.80 to 0.90, and a few are a little larger than one.

Table 3.4 presents the parameters-assignment results for the mixture of two gamma GB2 model. It provides the MH sampler acceptance rate for parameters, Geweke convergence diagnostic test, and effective sample sizes. The acceptance rate for parameters $(\alpha, \boldsymbol{\beta})$ for the models without random area effects and parameters $(\alpha, \boldsymbol{\beta}, \sigma^2)$ for the models with random area effects are provided. This table has every p-value greater than 0.05. The effective sample sizes for all parameters are unity except very few have less, about 0.80 to 0.90, and a few are a little larger than unity.

Table 3.5 presents the parameters-assignment results for the mixture of two generalized gamma GB2 model. It provides the MH sampler acceptance rate for parameters, Geweke convergence diagnostic test, and effective sample sizes. The acceptance rate for parameters $(\alpha, \gamma, \boldsymbol{\beta})$ for models without random area effects and parameters $(\alpha, \gamma, \boldsymbol{\beta}, \sigma^2)$ for models with random area effects are provided. This table has every p-value greater than 0.05. The effective sample sizes for all parameters are unity except very few have less, about 0.80 to 0.90, and a few are a little larger than unity.

Now, we show the trace and correlation plots. The trace plots for parameters sampled from the mixture of two generalized gamma (GB2) model with random area effects are shown for the *Mountains stratum* (stratum one). The trace plots for parameters alpha, gamma, sigma square, and vector of beta coefficients are shown from Figure 3.1 to Figure 3.13. The correlation plot for parameters sampled from the mixture of two generalized gamma (GB2) model with random area effects is shown for *Mountains stratum* (stratum one). The correlation plots for parameters alpha, gamma, sigma square, and vector of beta coefficients are shown from Figure 3.14 to Figure 3.26. The trace and correlation plots for other strata are also similar and not shown here.

Below we discuss the plots of the densities and mean responses in the PSUs plots. We have provided plots from the mixture of two generalized gamma (GB2) model with random area effects, the selected model from LPML criterion assuming noisy responses. The density plots of observed responses and overlaying predicted responses for all stratum are provided. The diagonal plots for observed mean responses versus predicted mean responses in the PSUs are provided for all stratum.

Figure 3.27 shows the overlaying density plot of the observed responses variable and a set of 1,000 predicted responses by the mixture of two generalized gamma GB2 model with random area effects for the *Mountains stratum* (stratum one). The black line is for an observed response variable and red lines are for the predicted responses. Figure 3.28 shows the diagonal plot for comparing mean responses in the PSUs for observed and predicted responses by the mixture of two generalized gamma GB2 model with random area effects in the *Mountains stratum* (stratum one). We present similar figures for all other strata from the generalized gamma models with random area effects.

Figures 3.29 and 3.30 are overlaying density plots of the observed welfare response variable and predicted responses, and a diagonal plot for mean responses in the PSUs for observed and predicted responses for stratum 2 *(Kathmandu valley urban areas)*. Figures 3.31 and 3.32 are overlaing density plots of the observed welfare response variable and predicted responses, and a diagonal plot for mean responses in the PSUs for observed and predicted responses for stratum 3 *(Other hills urban areas)*. Figures 3.33 and 3.34 are overlaying density plots of the observed welfare response variable and predicted responses, and a diagonal plot for mean responses in the PSUs for observed and predicted responses for stratum 4 *(Hill rural areas)*. Figures 3.35 and 3.36 are overlaying density plots of the observed welfare response variable and predicted responses, and a diagonal plot for mean responses in the PSUs for observed and predicted responses for stratum 5 *(Terai urban areas)*. Figures 3.37 and 3.38 are overlaying density plots of the observed welfare response variable and predicted responses, and a diagonal plot for mean responses in the PSUs for observed and predicted responses for stratum 6 *(Terai rural areas)*.

Finally, we note that a formal article is under preparation on the topic *"Hierarchical Bayesian models for noisy size responses from small areas: An application to poverty estimation"* (Manandhar and Nandram, 2017b).

Table 3.1: LPML values for three GB2 models
(with and without random area effects)

Models without random area effects

Stratum	Model		
	Expo-Gamma	**Gamma-Gamma**	**GGgamma-GGgamma**
1	-477.3	-221.8	-220.1
2	-728.7	-623.8	-620.5
3	-498.5	-391.3	-386.8
4	-1454.7	-896.4	-895.9
5	-519.1	-377.2	-376.6
6	-1729.3	-1090.5	-1090.3

Models with random area effects

Stratum	Model			
	HB NER	**Expo-Gamma**	**Gamma-Gamma**	**GGgamma-GGgamma**
1	-218.4	-487.6	-195.2	-173.9
2	-608.0	-741.5	-614.7	-591.8
3	-375.6	-509.6	-383.2	-362.1
4	-768.6	-1479.3	-809.4	-756.3
5	-357.9	-530.8	-372.5	-349.0
6	-1022.0	-1766.2	-1063.5	-1008.0

Table 3.2: Percent of CPO values below 0.02 for three GB2 models
(with and without random area effects)

Models without random area effects

Stratum	Model		
	Expo-Gamma	**Gamma-Gamma**	**GGgamma-GGgamma**
1	0.00	2.60	2.60
2	1.96	3.92	4.17
3	2.08	2.98	2.98
4	0.52	2.34	2.34
5	1.23	3.92	3.92
6	0.65	2.53	2.53

Models with random area effects

Stratum	Model			
	HB NER	**Expo-Gamma**	**Gamma-Gamma**	**GGgamma-GGgamma**
1	2.08	0.26	1.56	1.30
2	3.68	2.21	3.69	6.62
3	2.38	1.79	2.08	2.68
4	1.82	0.35	1.39	1.48
5	2.94	1.47	2.70	2.45
6	2.53	0.74	2.53	2.37

Table 3.3: Exponential-Exponential mixture GB2 models: Metropolis–Hastings acceptance rates, Geweke test p-values and effective sample sizes

Models without area effects

Stratum	Beta acceptance rate	P-values									
		Beta0	**Beta1**	**Beta2**	**Beta3**	**Beta4**	**Beta5**	**Beta6**	**Beta7**	**Beta8**	**Beta9**
1	0.25	0.60	0.54	0.57	0.38	0.62	0.81	0.57	0.67	0.83	0.42
2	0.21	0.79	0.09	0.40	0.50	0.53	0.51	0.92	0.42	0.15	0.76
3	0.25	0.93	0.40	0.86	0.57	0.54	0.97	0.43	0.33	0.02	0.56
4	0.25	0.33	0.21	0.18	0.80	0.30	0.67	0.11	0.83	0.19	0.91
5	0.25	0.54	0.20	0.42	0.41	0.93	0.85	0.10	0.49	0.72	0.19
6	0.26	0.23	0.92	0.47	0.67	0.75	0.13	0.55	0.38	0.99	0.33

Stratum	Effective sample size									
	Beta0	**Beta1**	**Beta2**	**Beta3**	**Beta4**	**Beta5**	**Beta6**	**Beta7**	**Beta8**	**Beta9**
1	1.00	1.00	1.00	1.00	1.00	1.00	0.80	1.00	1.13	1.00
2	1.00	1.00	1.00	1.00	1.00	0.74	0.94	1.00	1.00	1.00
3	1.00	1.00	1.00	0.87	1.00	0.80	1.00	1.00	0.89	1.00
4	1.00	0.87	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
5	1.15	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.97
6	0.90	1.20	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Models with area effects

Stratum	Sigma2 Beta acceptance rate	P-values										
		Sigma	**Beta0**	**Beta1**	**Beta2**	**Beta3**	**Beta4**	**Beta5**	**Beta6**	**Beta7**	**Beta8**	**Beta9**
1	0.524	0.38	0.65	0.97	0.57	0.21	0.32	0.97	0.84	0.64	0.39	0.95
2	0.524	0.23	0.85	0.70	0.80	0.24	0.09	0.80	0.65	0.75	0.30	0.77
3	0.533	0.77	0.15	0.13	0.73	0.88	0.59	0.53	0.19	0.26	0.74	0.28
4	0.509	0.27	0.94	0.33	0.76	0.33	0.91	0.61	0.28	0.36	0.67	0.17
5	0.518	0.71	0.36	0.51	0.31	0.81	0.37	0.94	0.49	0.24	0.97	0.70
6	0.515	0.18	0.42	0.22	0.93	0.91	0.92	0.48	0.35	0.71	0.28	0.85

Stratum	Effective sample size										
	Sigma	**Beta0**	**Beta1**	**Beta2**	**Beta3**	**Beta4**	**Beta5**	**Beta6**	**Beta7**	**Beta8**	**Beta9**
1	1.00	1.24	1.00	1.00	0.87	1.00	1.00	1.43	1.23	1.00	1.00
2	1.00	1.00	1.00	1.00	0.86	1.00	1.00	1.00	1.13	1.00	1.00
3	1.00	1.13	1.11	1.03	1.00	1.00	1.10	1.24	1.00	1.00	1.00
4	1.00	1.00	1.00	1.00	1.00	1.65	1.00	1.00	1.00	1.17	1.00
5	1.14	1.00	1.00	1.00	1.32	1.00	1.17	1.31	1.00	1.19	1.00
6	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.89	1.00	1.00	1.32

Table 3.4: Gamma-Gamma mixture GB2 models: Metropolis–Hastings acceptance rates, Geweke test p-values and effective sample sizes

Models without area effects

Stratum	**Alpha Beta acceptance rate**	**P-values**										
		Alpha	**Beta0**	**Beta1**	**Beta2**	**Beta3**	**Beta4**	**Beta5**	**Beta6**	**Beta7**	**Beta8**	**Beta9**
1	0.475	0.46	0.09	0.25	0.78	0.64	0.30	0.35	0.52	0.09	0.09	0.40
2	0.559	0.26	0.41	0.24	0.47	0.18	0.86	0.57	0.55	0.86	0.50	0.37
3	0.543	0.89	0.51	0.30	0.12	0.30	0.17	0.62	0.88	0.29	0.46	0.19
4	0.538	0.32	0.14	0.75	0.46	0.66	0.70	0.33	0.60	0.32	0.40	0.36
5	0.561	0.91	0.83	0.32	0.52	0.97	0.12	0.37	0.65	0.74	0.31	0.90
6	0.554	0.81	0.92	0.25	0.56	0.65	0.41	0.28	0.09	0.58	0.37	0.79

Stratum	**Effective sample size**										
	Alpha	**Beta0**	**Beta1**	**Beta2**	**Beta3**	**Beta4**	**Beta5**	**Beta6**	**Beta7**	**Beta8**	**Beta9**
1	1.00	1.00	1.00	0.88	1.00	1.00	1.00	1.00	1.00	1.08	2.50
2	1.34	1.08	1.00	1.00	1.00	1.00	1.12	1.00	1.00	1.18	1.13
3	0.94	0.86	1.00	1.00	1.09	1.00	1.00	1.00	1.00	1.00	1.00
4	1.00	1.00	1.11	1.39	1.00	0.91	1.16	1.09	0.86	1.00	1.00
5	1.00	1.00	1.00	1.00	1.00	0.87	1.00	1.50	1.00	1.13	1.31
6	1.00	1.00	1.00	1.00	1.00	0.92	0.90	1.00	1.25	0.79	1.00

Models with area effects

Stratum	**Alpha, Beta, Sigma2 acceptance rate**	**P-values**											
		Alpha	**Sigma2**	**Beta0**	**Beta1**	**Beta2**	**Beta3**	**Beta4**	**Beta5**	**Beta6**	**Beta7**	**Beta8**	**Beta9**
1	0.529	0.85	0.36	0.35	0.73	0.94	0.78	0.55	0.11	0.49	0.21	0.41	0.41
2	0.524	0.51	0.85	0.32	0.94	0.14	0.17	0.71	0.22	0.79	0.58	0.36	0.14
3	0.533	0.11	0.44	0.25	0.33	0.16	0.87	0.56	0.66	0.44	0.74	0.44	0.22
4	0.512	0.73	0.22	0.67	0.62	0.24	0.88	0.82	0.86	0.63	0.57	0.52	0.78
5	0.537	0.20	0.23	0.12	0.42	0.14	0.84	0.63	0.60	0.53	0.69	0.15	0.68
6	0.508	0.15	0.33	0.18	0.97	0.37	0.27	0.44	0.58	0.07	0.23	0.63	0.11

Stratum	**Effective sample size**											
	Alpha	**Sigma2**	**Beta0**	**Beta1**	**Beta2**	**Beta3**	**Beta4**	**Beta5**	**Beta6**	**Beta7**	**Beta8**	**Beta9**
1	0.92	1.21	1.00	1.00	1.04	1.00	1.11	1.00	1.00	1.00	1.00	1.00
2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.90	1.00	1.00	1.00	1.00
3	1.09	1.00	1.00	1.00	1.00	1.00	1.21	0.85	1.14	1.00	1.00	1.00
4	1.24	1.00	1.00	1.00	1.00	1.61	1.00	1.00	0.88	1.00	1.00	1.00
5	1.59	1.00	1.33	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
6	1.00	1.07	1.00	1.00	1.00	1.13	1.00	1.00	1.10	1.00	1.00	1.00

Table 3.5: Generalized Gamma-Generalized Gamma mixture GB2 models: Metropolis–Hastings acceptance rates, Geweke test p-values and effective sample sizes

Models without area effects

Stratum	**Alpha, Gamma, Beta acceptance rate**	**P-values**											
		Alpha	**Gamma**	**Beta0**	**Beta1**	**Beta2**	**Beta3**	**Beta4**	**Beta5**	**Beta6**	**Beta7**	**Beta8**	**Beta9**
1	0.545	0.97	0.42	0.28	0.71	0.53	0.96	0.09	0.36	0.14	0.28	0.59	0.38
2	0.547	0.58	0.80	0.42	0.53	0.42	0.40	0.39	0.64	0.26	0.40	0.15	0.79
3	0.583	0.80	0.50	0.38	0.90	0.69	0.81	0.99	0.57	0.62	0.18	0.85	0.61
4	0.554	0.31	0.41	0.31	0.87	0.16	0.17	0.91	0.27	0.39	0.43	0.36	0.15
5	0.552	0.35	0.62	0.51	0.39	0.85	0.18	0.62	0.27	0.47	0.49	0.24	0.16
6	0.563	0.39	0.34	0.56	0.68	0.11	0.62	0.81	0.15	0.93	0.08	0.19	0.67

Stratum	**Effective sample size**											
	Alpha	**Gamma**	**Beta0**	**Beta1**	**Beta2**	**Beta3**	**Beta4**	**Beta5**	**Beta6**	**Beta7**	**Beta8**	**Beta9**
1	1.00	1.00	1.00	1.00	1.11	1.00	1.00	1.00	1.00	1.00	1.00	0.85
2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.14	1.00
3	0.91	1.00	1.00	0.89	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
4	1.00	1.00	1.00	1.00	1.12	1.02	1.00	1.00	1.23	1.00	0.82	1.00
5	1.00	1.00	1.00	1.00	0.91	1.00	1.00	1.00	1.00	1.00	1.00	1.00
6	1.00	1.00	0.72	1.13	0.68	1.00	1.00	1.09	0.90	0.84	1.00	1.24

Models with area effects

Stratum	**Alpha, Gamma, Beta, Sigma2 acceptance rate**	**P-values**												
		Alpha	**Gamma**	**Sigma2**	**Beta0**	**Beta1**	**Beta2**	**Beta3**	**Beta4**	**Beta5**	**Beta6**	**Beta7**	**Beta8**	**Beta9**
1	0.526	0.11	0.10	0.61	0.26	0.19	0.11	0.06	0.90	0.66	0.39	0.18	0.14	0.34
2	0.547	0.30	0.22	0.27	0.24	0.79	0.48	0.16	0.62	0.24	0.18	0.37	0.29	0.25
3	0.533	0.81	0.99	0.46	0.18	0.68	0.45	0.64	0.56	0.98	0.38	0.18	0.35	0.60
4	0.513	0.84	0.90	0.60	0.83	0.05	0.08	0.70	0.02	0.81	0.07	0.15	0.69	0.89
5	0.545	0.58	0.80	0.49	0.42	0.91	0.34	0.36	0.09	0.61	0.38	0.79	0.73	0.61
6	0.501	0.32	0.23	0.97	0.66	0.71	0.49	0.78	0.92	0.10	0.14	0.05	0.14	0.56

Stratum	**Effective sample size**												
	Alpha	**Gamma**	**Sigma2**	**Beta0**	**Beta1**	**Beta2**	**Beta3**	**Beta4**	**Beta5**	**Beta6**	**Beta7**	**Beta8**	**Beta9**
1	1.00	1.00	1.10	1.00	1.00	1.00	1.00	1.12	1.00	1.00	1.00	1.33	1.00
2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.10	1.00	1.09	1.11	1.00	1.00
3	0.79	0.77	1.00	1.00	1.00	1.00	0.85	1.00	1.00	1.00	1.00	1.00	1.00
4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.15	1.00	1.12	1.03
5	0.84	0.83	1.00	1.00	1.00	1.00	1.00	1.00	0.88	0.93	1.00	1.00	1.00
6	1.00	1.00	1.00	1.14	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.37	0.91

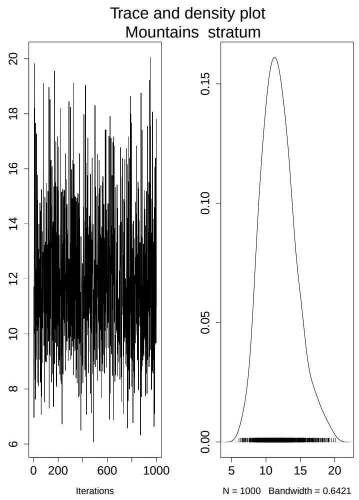


Figure 3.1: Alpha

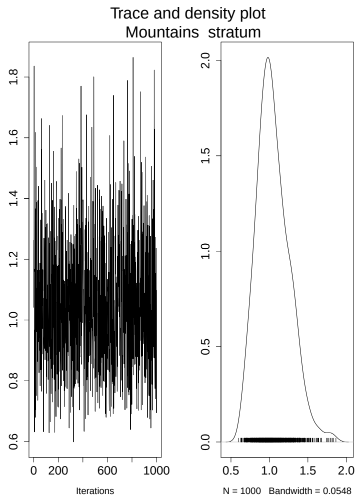


Figure 3.2: Gamma

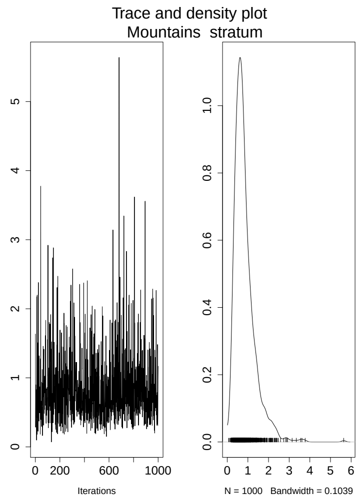


Figure 3.3: Sigma Square

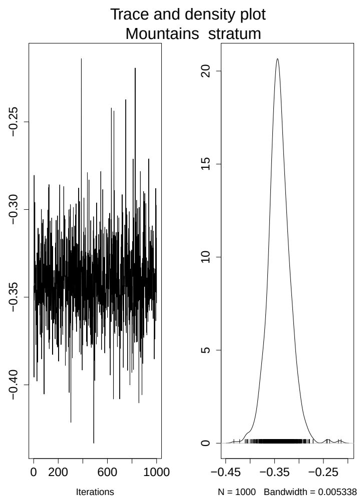


Figure 3.4: Beta0

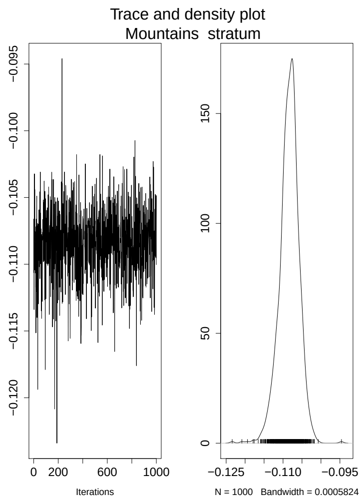


Figure 3.5: Beta1

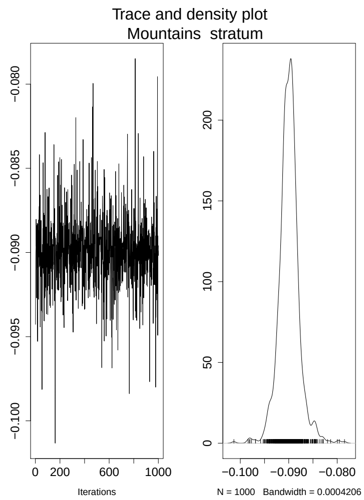


Figure 3.6: Beta2

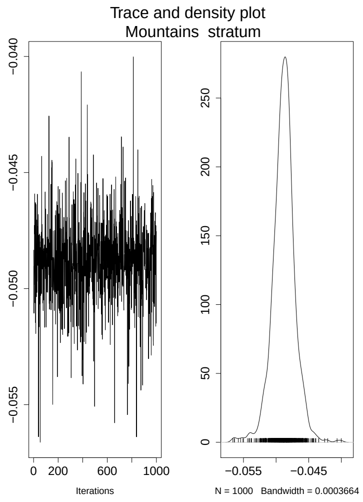


Figure 3.7: Beta3

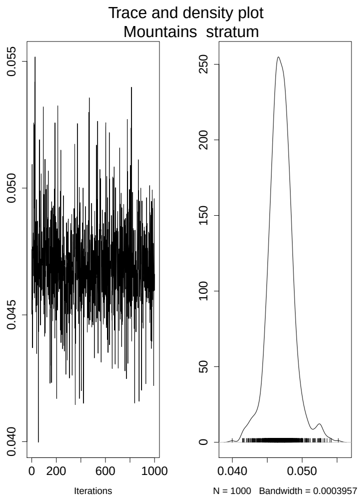


Figure 3.8: Beta4

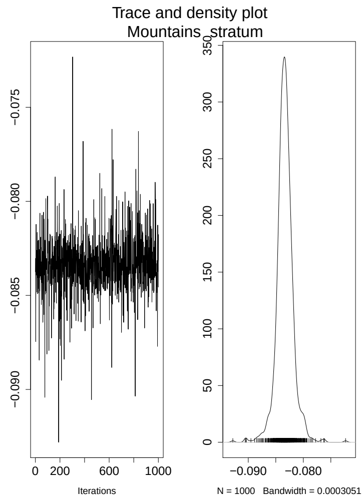


Figure 3.9: Beta5

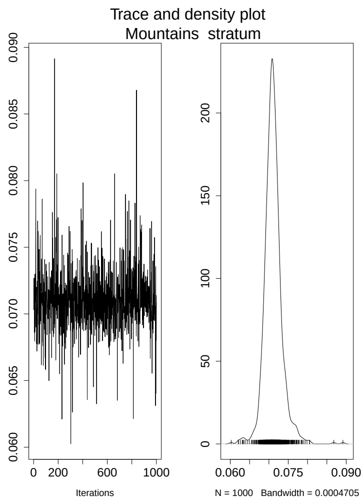


Figure 3.10: Beta6

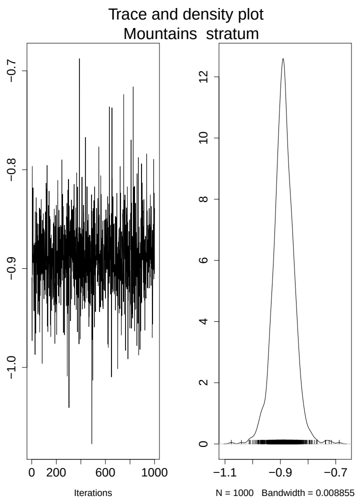


Figure 3.11: Beta7

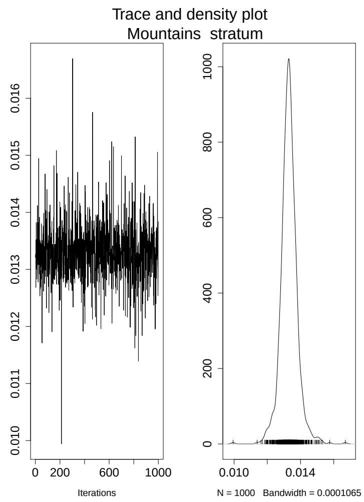


Figure 3.12: Beta8

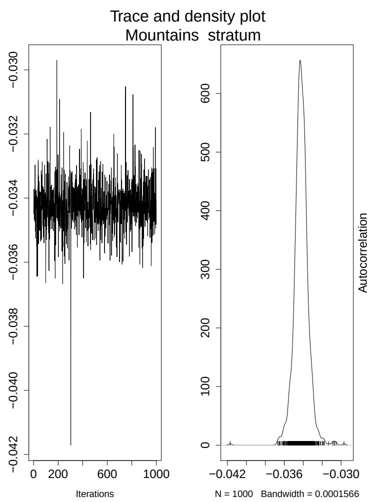


Figure 3.13: Beta9

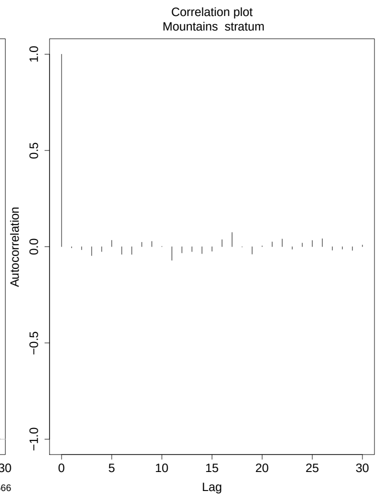


Figure 3.14: Alpha

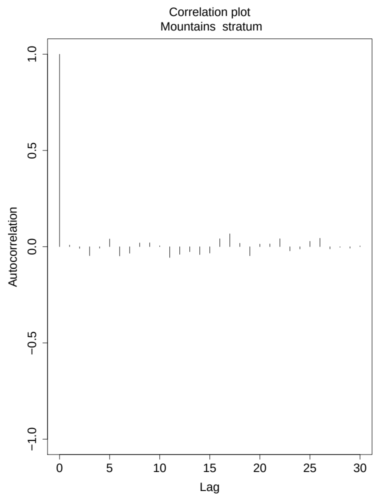


Figure 3.15: Gamma

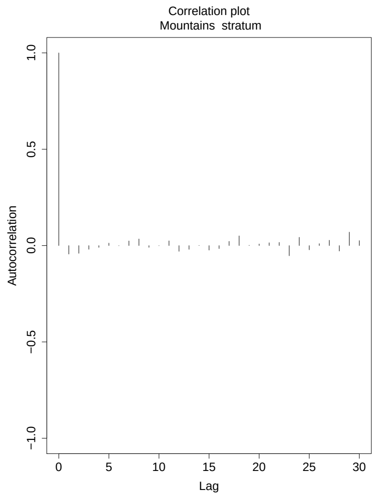


Figure 3.16: Sigma Square

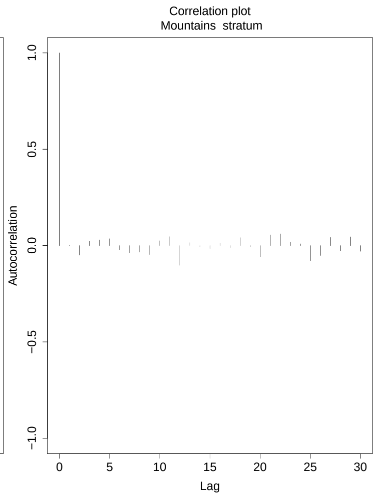


Figure 3.17: Beta0

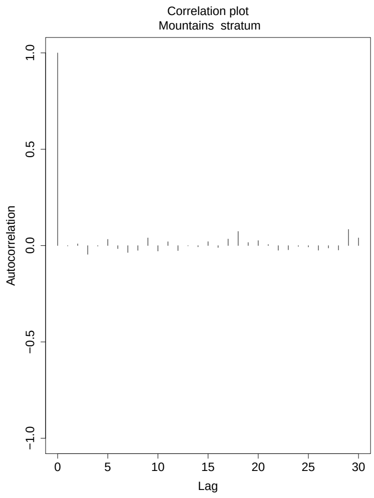


Figure 3.18: Beta1

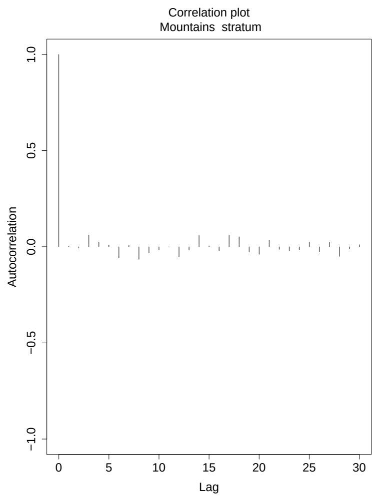


Figure 3.19: Beta2

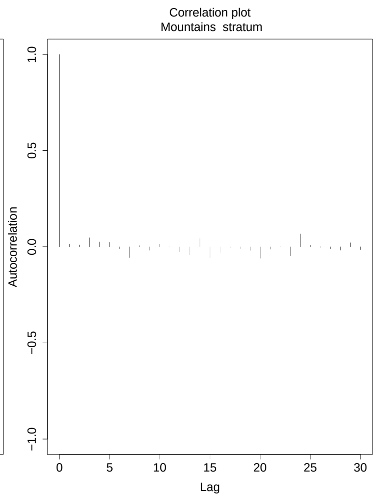


Figure 3.20: Beta3

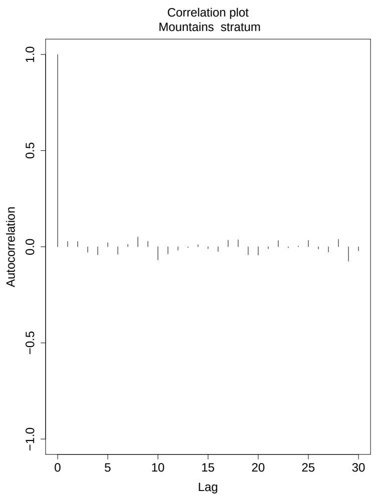


Figure 3.21: Beta4

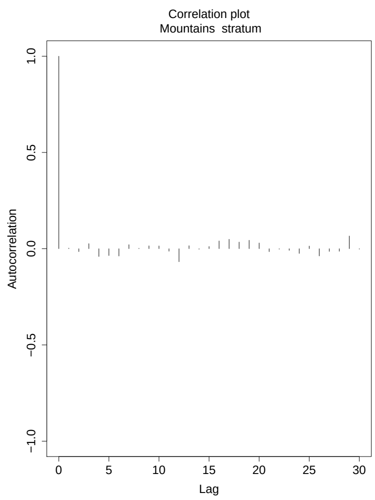


Figure 3.22: Beta5

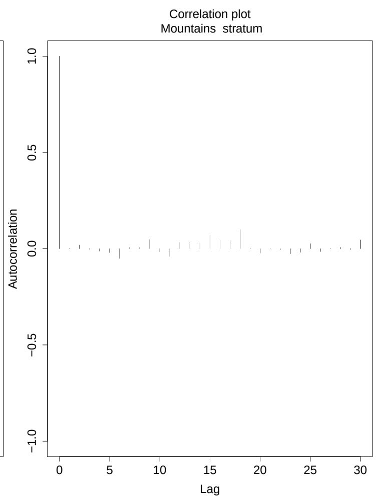


Figure 3.23: Beta6

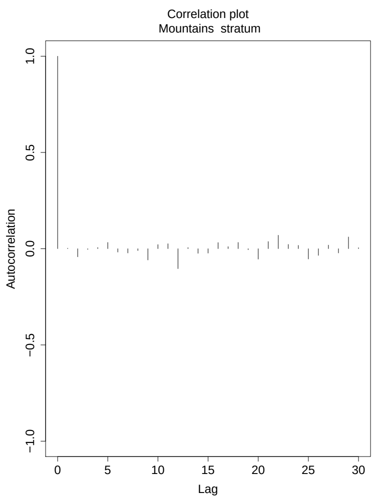


Figure 3.24: Beta7

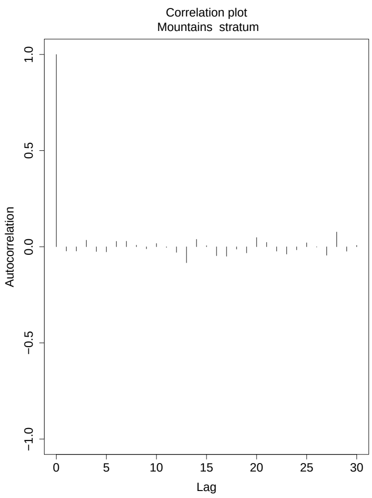


Figure 3.25: Beta8

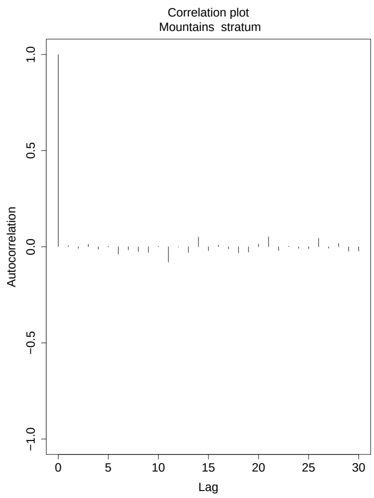


Figure 3.26: Beta9

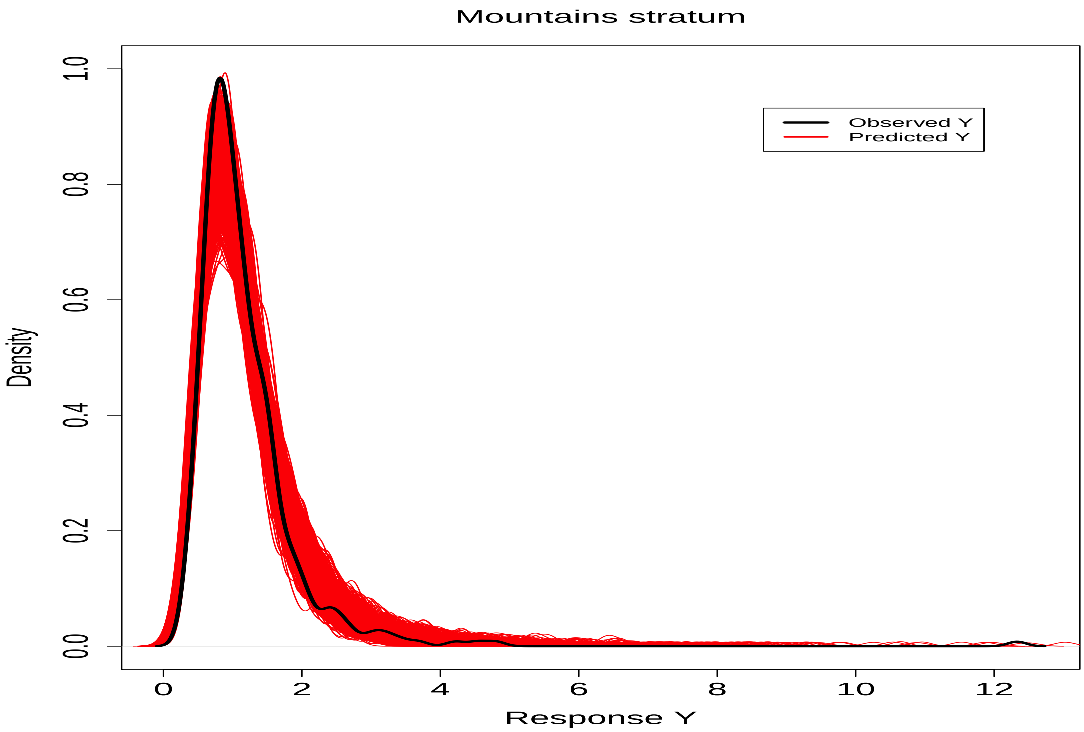


Figure 3.27: Observed and predicted responses density plots

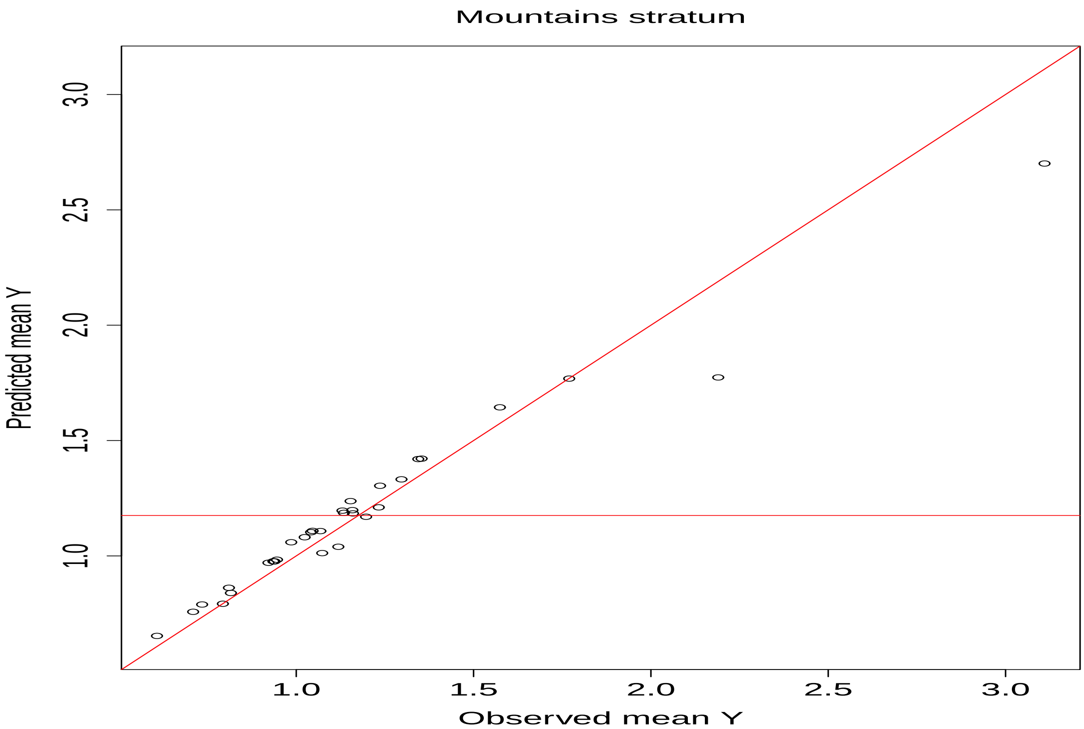


Figure 3.28: Observed and predicted mean responses by PSU

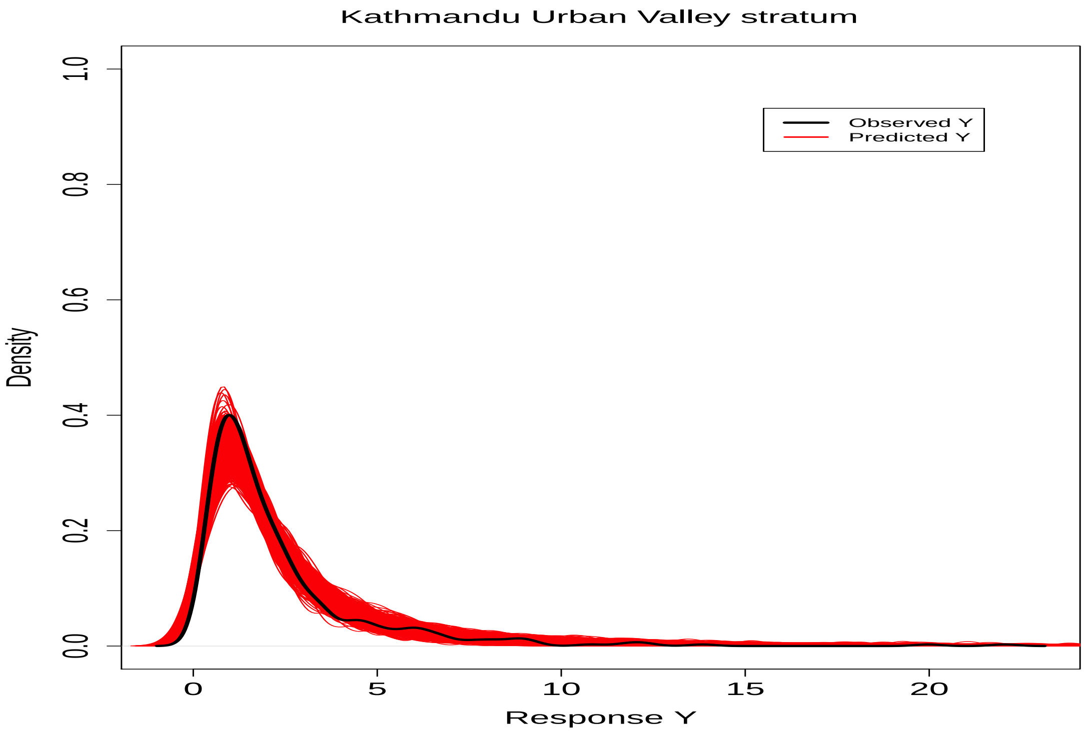


Figure 3.29: Observed and predicted responses density plots

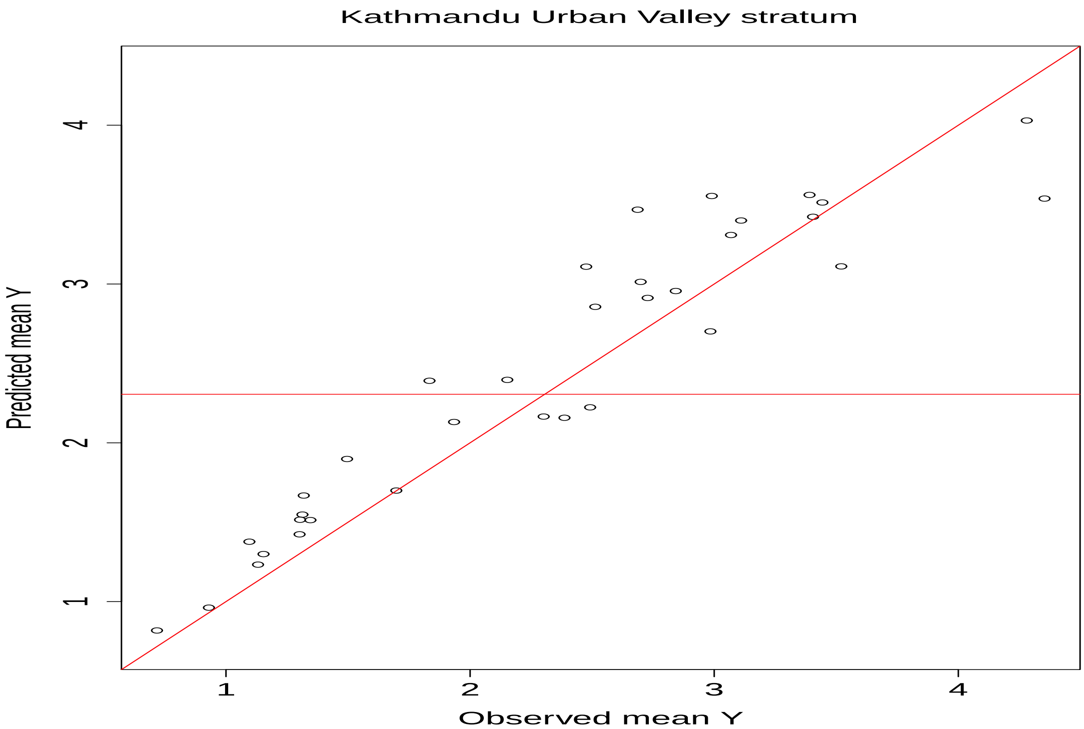


Figure 3.30: Observed and predicted mean responses by PSU

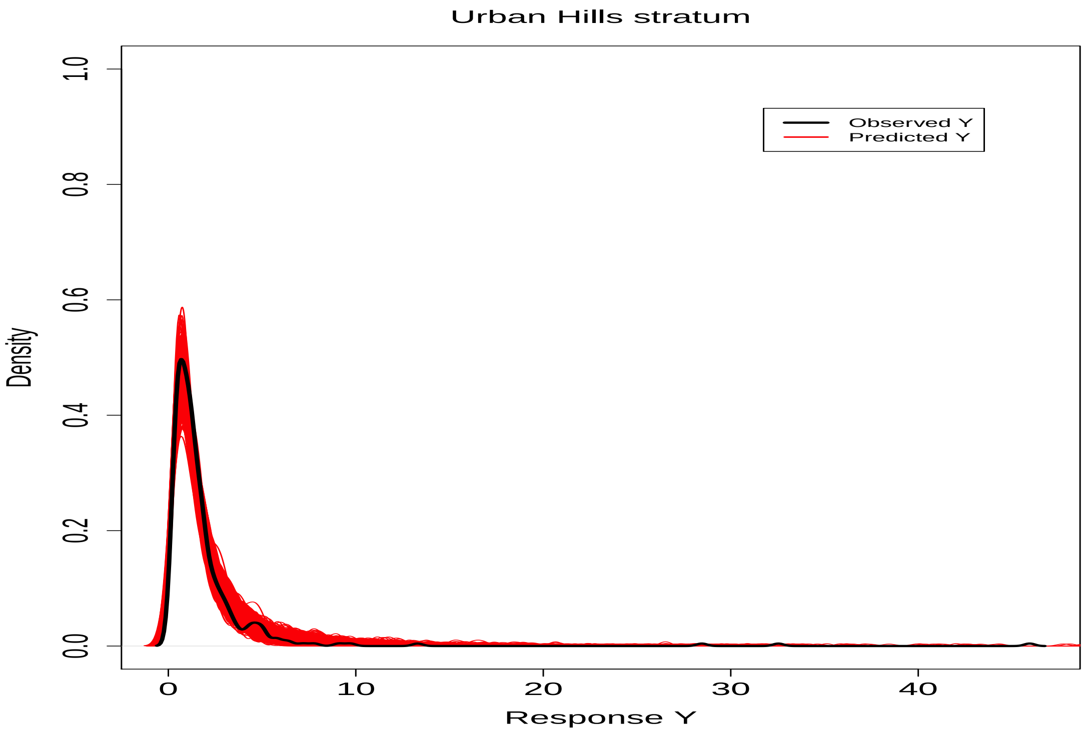


Figure 3.31: Observed and predicted responses density plots

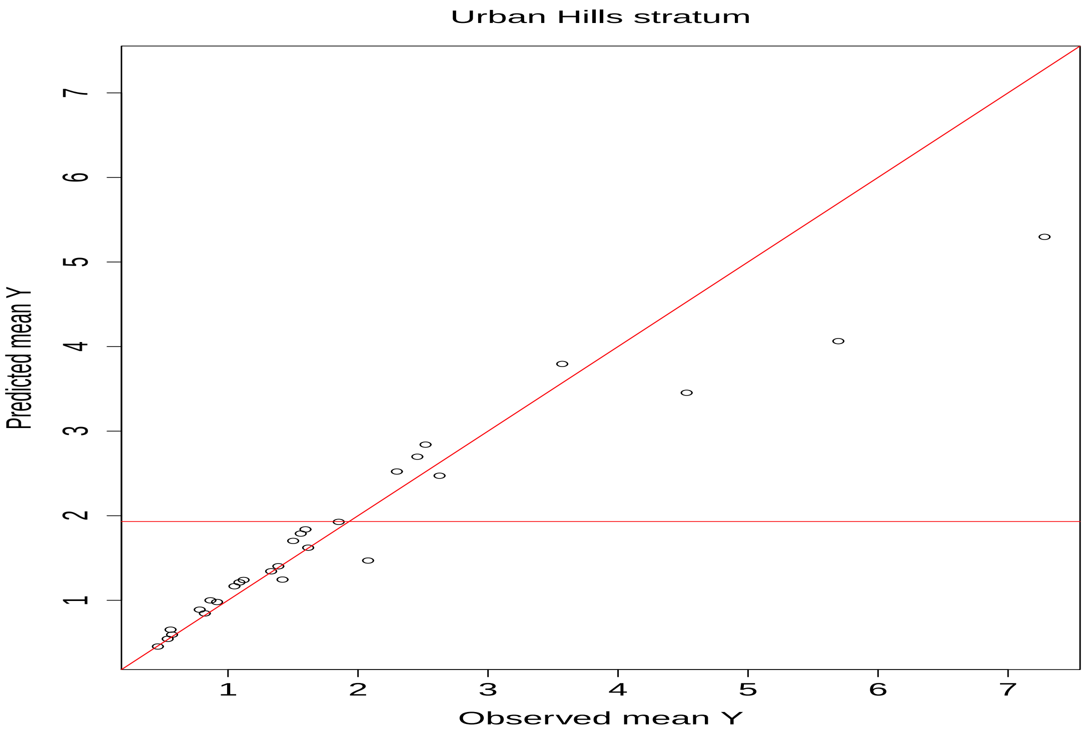


Figure 3.32: Observed and predicted mean responses by PSU

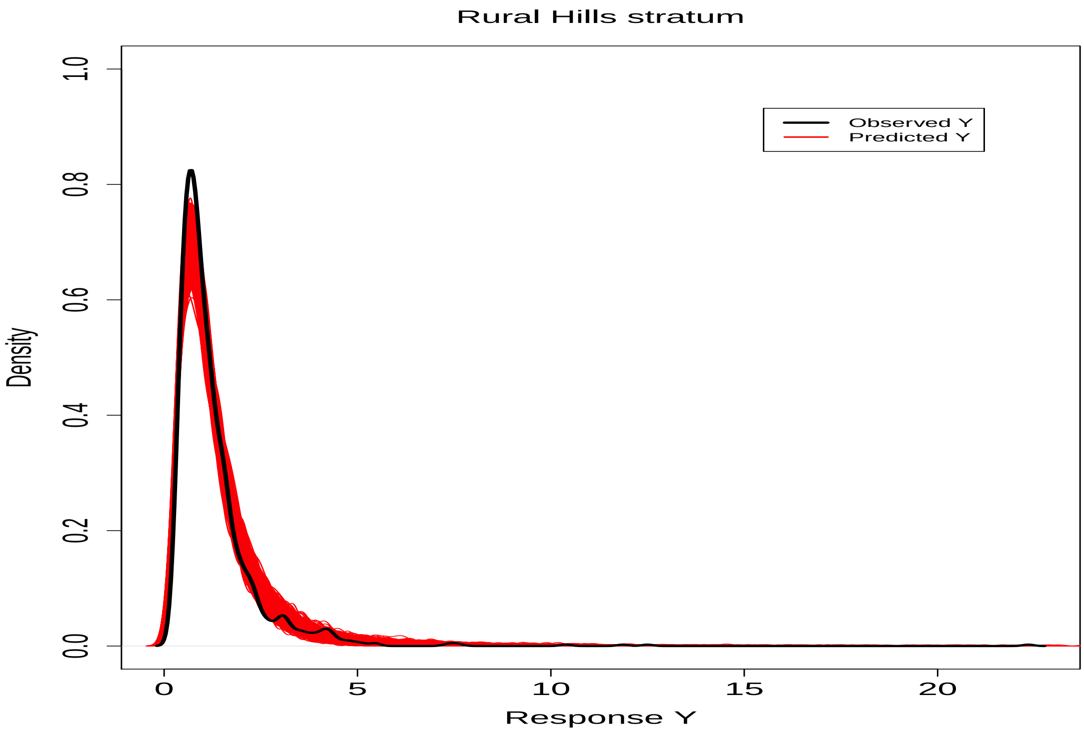


Figure 3.33: Observed and predicted responses density plots

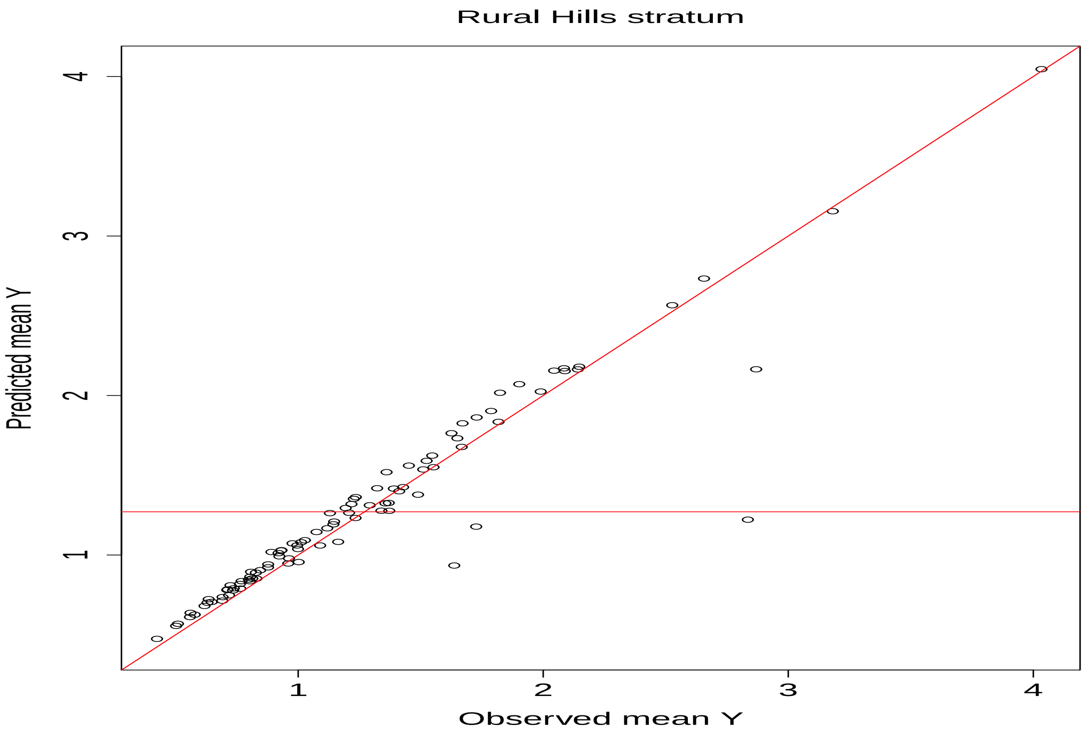


Figure 3.34: Observed and predicted mean responses by PSU

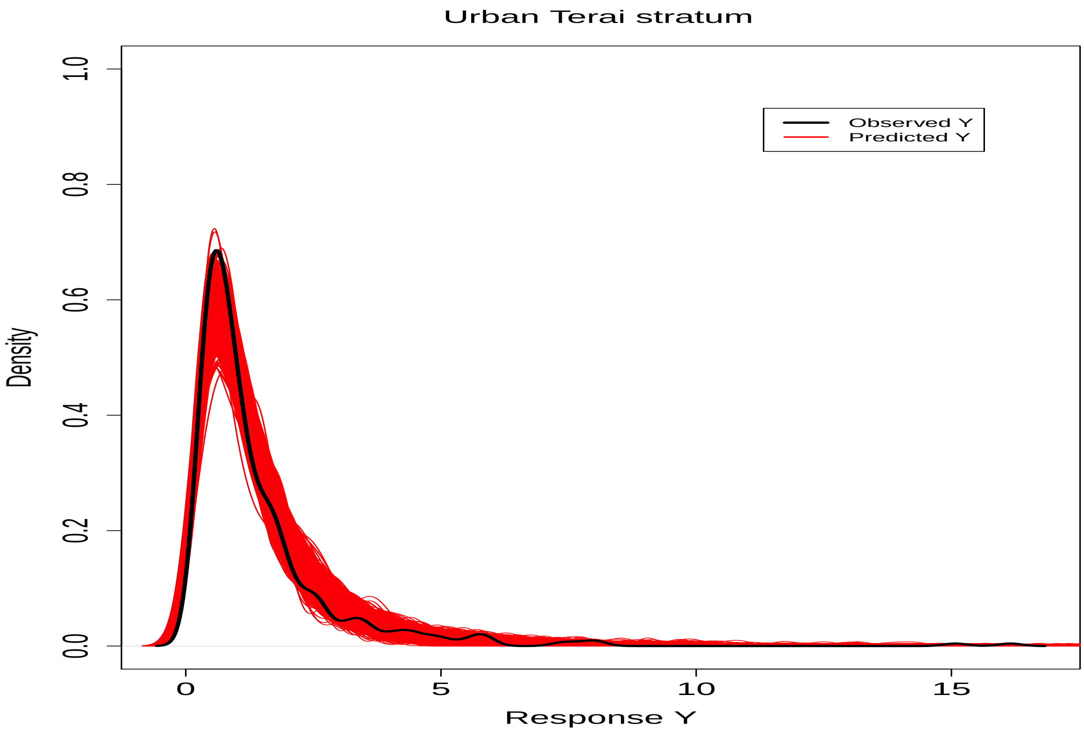


Figure 3.35: Observed and predicted responses density plots

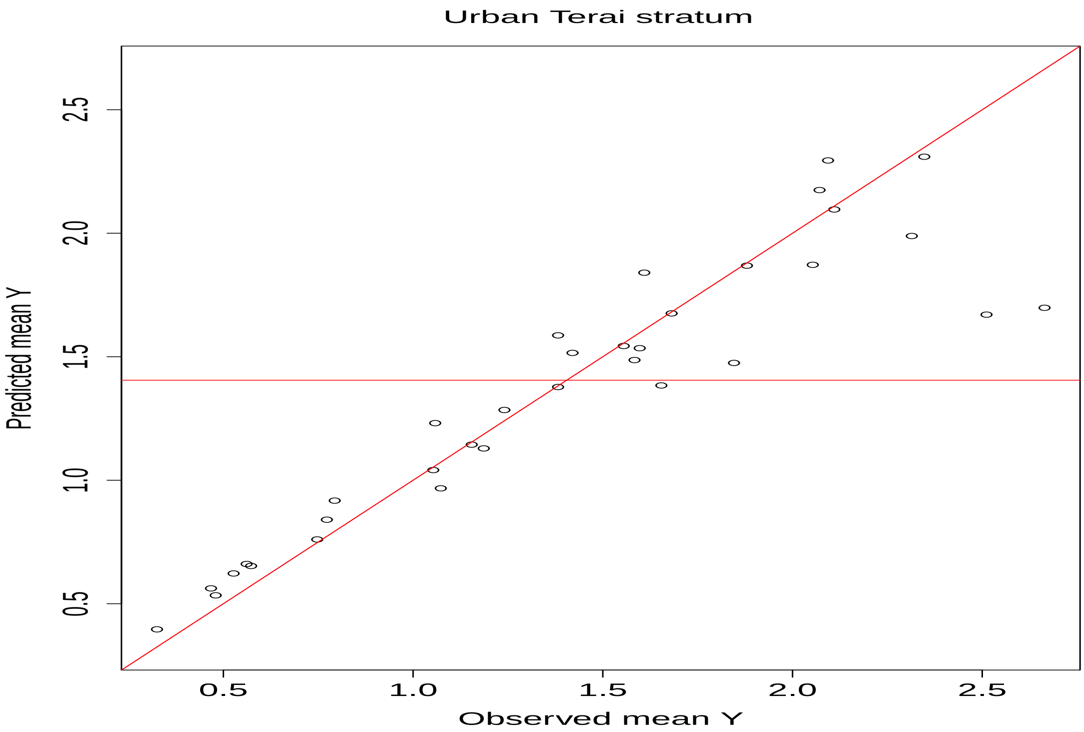


Figure 3.36: Observed and predicted mean responses by PSU

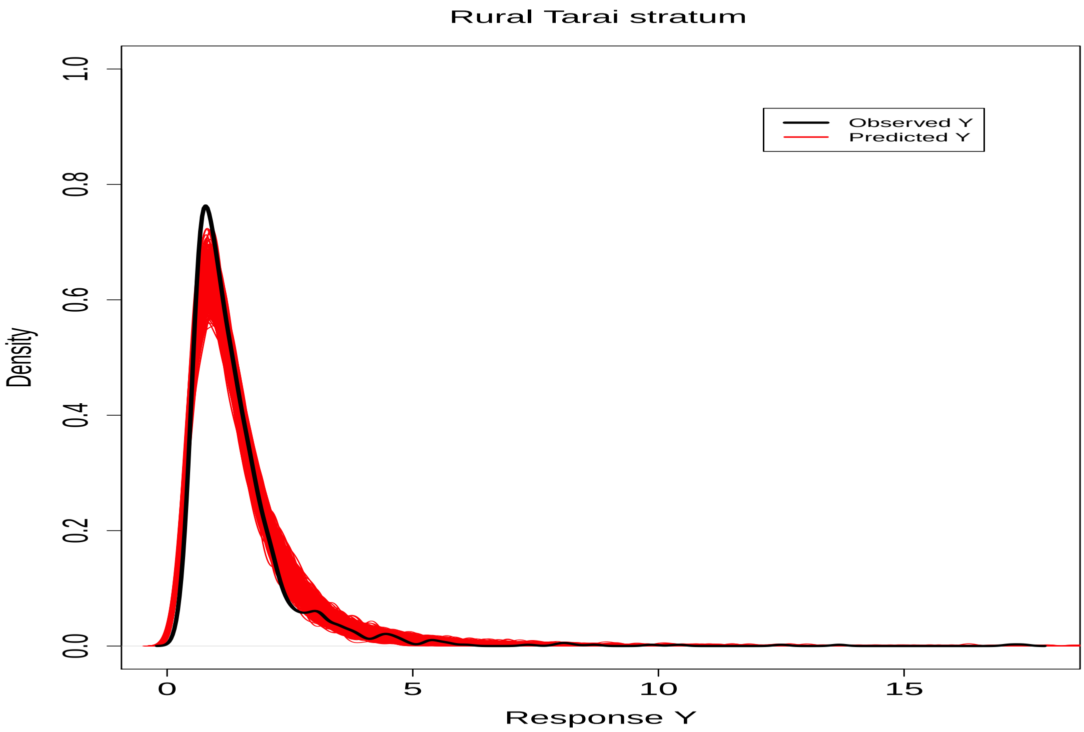


Figure 3.37: Observed and predicted responses density plots

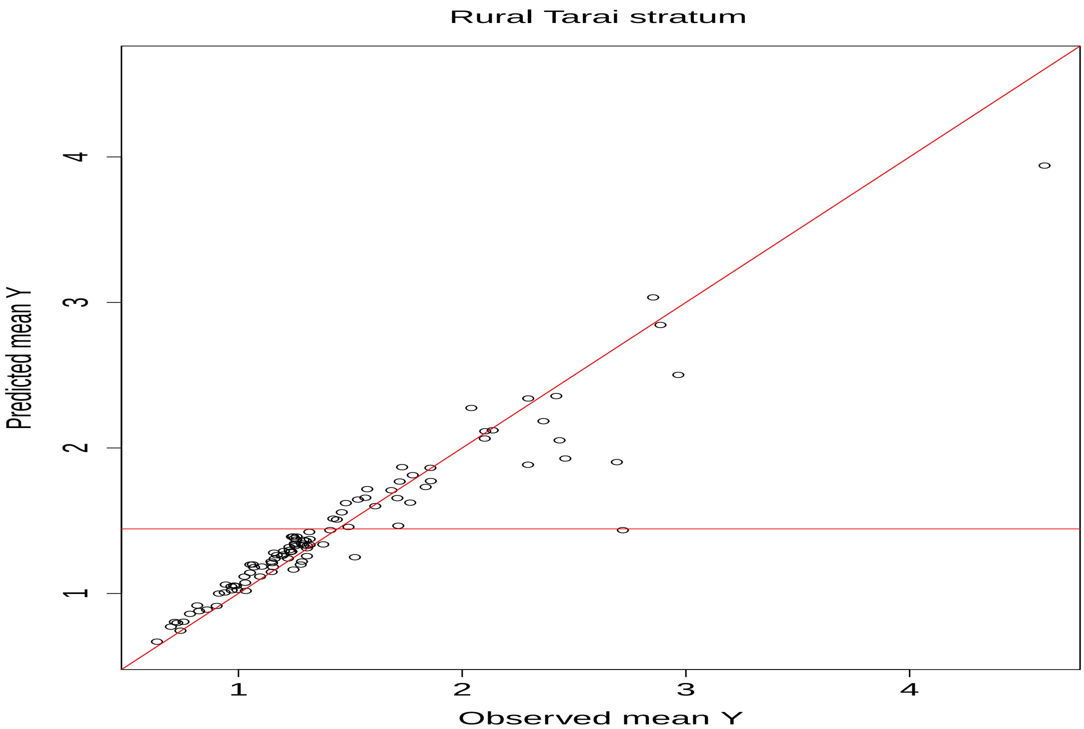


Figure 3.38: Observed and predicted mean responses by PSU

Chapter 4

Finite Population Quantities, Simulation and Conclusion

In this chapter we apply the selected hierarchical Bayesian models to predict the responses in the large survey, population census data. We provide simulation study of the selected models to support our models. Finally, we conclude by stating our contributions in methodology and future works.

4.1 Census Data and Small Area Estimation

We have developed models for continuous and positively skewed (CPS) size data from small areas. We have used exponential to GB2 densities, support $(0, \infty)$, to develop hierarchical Bayesian models. We have applied models to the welfare per capita consumption data with nine covariates from NLSS-II, 2003/04. We have these nine covariates generated in the census 2001 data, for the purpose of SAE. In Chapter 2, we have considered noiseless responses and fit standard distributions: the exponential, the gamma and the generalized gamma. In Chapter 3, we have considered noisy responses and fit GB2 distributions. We have defined noisy responses as recalling errors introduced in the responses. More information about response and possible noises is included in subsection *"Response variable"*, section *"Application"*, Chapter 1. The model assignment tables and figures are presented in their respective chapters.

In this chapter, we will apply the selected models for SAE using 2001 population census data. Table 4.1 shows the distribution of political divisions in Nepal. Note that these political boundaries have not been changed since the period of the Panchyet dictator monarchy system. It has 5 development regions, 14 zones, 75 districts, 3,972 VDC/municipalities and 36,032 wards at the time of the 2001 population census and the NLSS-II survey, 2003/04. As discussed in Chapter 1, NLSS-II is a very small scale survey with only 3,912 households enumerated in a total of 4,311,747 households in the sample frame. That is, 326 PSUs are selected in NLSS-II from a total of 36,067 PSUs, Table 1.1. Owing to the small sample sizes, this survey can give reliable estimates only at stratum level or higher but cannot give estimates for small areas like districts, municipality/VDCs or wards.

Table 4.1: Political divisions in Nepal

Political divisions	**Count**
Regions	5
Zones	14
Districts	75
VDC/Municipalities	3972
Wards	36032

We have the best fitted models via the LPML criterion for noiseless and noisy responses for small area estimation. For *noiseless responses* LPML values suggest the *generalized gamma model with random area effects*, and for *noisy responses* LPML values suggest the *mixture of two generalized gamma (GB2) model with random area effects*.

To facilitate SAE, we have nine covariates both in the NLSS-II and the population census and their consistencies were checked prior to use. Poverty indicators have been calculated using the poverty threshold of an average of 7,696 Nepalese rupees per year in 2003, adjusted for spatial price variation as reported in NLSS-II documents. It is the same poverty threshold used in SAE of Poverty, Nepal (Haslett et al., 2006).

We present the SAE of poverty indicators (poverty incidence, poverty gap, and poverty severity) by applying the generalized gamma and the mixture of two generalized gamma (GB2) models in the 2001 Population Census data. We give the district level estimates for *Mountains stratum* as an example. Estimates for all other strata can be calculated similarly and are not shown here. The small area estimations for VDC/Municipalities and for wards are also done in a similar way and not tabulated here.

4.2 Small Area Estimation

We have applied our models in population census data for SAE of poverty indicators. We have given the SAE of poverty incidences, poverty gaps, and poverty severities from two selected models. In this section we discuss results of the poverty indicators at the district level.

4.2.1 Selected models

Assuming *noiseless responses*, from Chapter 2 we have the *generalized gamma model with random area effects* as the best fitted model. We call this selected model the *noiseless model* from this point forward in this chapter for convenience. The *noiseless model* is

$$\begin{aligned} y_{ij}|\alpha, \boldsymbol{\beta}, \gamma, \nu_i &\overset{\text{ind}}{\sim} \text{GGamma}\left(\alpha, e^{-(\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\boldsymbol{\nu_i})}, \gamma\right), \quad \lambda_{ij} = e^{-(\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\boldsymbol{\nu_i})}, \\ \nu_i &\overset{\text{iid}}{\sim} N(0, \sigma^2), \quad i = 1, \cdots, \ell, \; j = 1, \cdots, n_i, \\ \pi(\boldsymbol{\beta}, \alpha, \sigma^2) &\propto \frac{1}{(1+\alpha^2)(1+\sigma^2)^2}, \\ \gamma &\sim \text{Gamma}(S, R), \quad \text{where shape `S' and rate `R' are specified.} \end{aligned}$$

Assuming *noisy responses*, from Chapter 3 we have the *mixture of two generalized gamma distributions (GB2) model with random area effects* as the best fitted model. We call this selected model the *noisy model* from this point forward in this chapter for convenience. The *noisy model* is

$$\begin{aligned} y_{ij}|\boldsymbol{\beta}, \alpha, \gamma, \nu_i &\overset{\text{ind}}{\sim} \text{GB2}\left(\alpha, e^{\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\boldsymbol{\nu_i}}, \gamma\right), \quad \theta_{ij} = e^{\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\boldsymbol{\nu_i}}, \; i = 1, \cdots, \ell, \; j = 1, \cdots, n_i, \\ \nu_i &\overset{\text{iid}}{\sim} N(0, \sigma^2), \\ \pi(\boldsymbol{\beta}, \alpha, \sigma^2) &\propto \frac{1}{(1+\alpha)^2 \, (1+\sigma^2)^2}, \\ \gamma &\sim \text{Gamma}(S, R), \quad \text{where shape `S' and rate `R' are specified.} \end{aligned}$$

4.2.2 Small Area Estimation from selected models

From each selected noiseless model and noisy models we have a set of 1,000 samples drawn for the parameters $(\alpha, \gamma, \boldsymbol{\beta}, \boldsymbol{\nu}, \sigma^2)$.

For SAE, first we get the area effects, $\nu_i, i = 1 \cdots, L$. In our model each PSU of sampled data has its own random area effects (parameters) drawn while developing the model. We note that, PSU in NLSS-II is a ward (the smallest political boundary) in most of the cases. However, in some cases it is a subset of a ward or the union of wards. Since PSU is created in the NLSS-II survey but the same PSU's geographical boundaries may or may not exist in the census data, to predict responses in the census data we consider the smallest political region (ward) as equivalent to the PSU in NLSS-II. Therefore, for prediction in the census data, we consider different random area effects parameters $\nu_i, i = 1. \cdots, L$ for each ward.

If PSUs are sampled, we already have information about random area effects $\nu_i, i = 1, \cdots, \ell$, and use that information for the ward to which it belongs. For the non-sampled PSUs, we do not have information about $\nu_i, i = \ell + 1, \cdots, L$. We draw the random area effects from the Bayesian bootstrap sampling method with prior Dirichlet($\mathbf{0}$) given all other parameters. See *Appendix B* for the Bayesian Bootstrap sampling procedure. Once we draw the random area effects, ν_i, we predict the responses in the population census data as follows:

Prediction Under Noiseless Model

(i) Find the rate parameters

$$\lambda_{ij} = e^{-(\boldsymbol{x}'_i \boldsymbol{\beta} + \nu_i)}.$$

(ii) Draw predicted responses from the generalized gamma distribution. Consider the transformation $t = y^\gamma$. This gives

$$\mathrm{G}_1 = (y_{ij})^\gamma \sim \text{Gamma}(\alpha, \lambda_{ij}), \quad \hat{y}_{ij} = \mathrm{G}_1^{\frac{1}{\gamma}}.$$

Prediction Under Noisy Model

(i) Find the rate parameters

$$\theta_{ij} = e^{\boldsymbol{x}'_{ij} \boldsymbol{\beta} + \nu_i}.$$

(ii) Draw shape parameter λ. In GB2 distribution consider a transformation $t = (\theta\lambda)^\gamma$. This gives $(\theta\lambda)^\gamma \sim$ Gamma $\left(\frac{\alpha+2}{\gamma}, 1\right)$. If we draw a random sample G_1 from this distribution, then we can calculate λ as

$$\mathrm{G}_1 = (\theta\lambda)^{\gamma} \sim \mathrm{Gamma}\left(\frac{\alpha+2}{\gamma}, 1\right), \quad \lambda_{ij} = \frac{\mathrm{G}_1^{\frac{1}{\gamma}}}{\theta_{ij}}.$$

(iii) Predict responses. In the GB2 distribution consider a transformation $t = (\lambda y)^{\gamma}$. This gives $(\lambda y)^{\gamma} \sim \mathrm{Gamma}\left(\frac{\alpha}{\gamma}, 1\right)$. If we draw a random sample G_2 from this distribution, then we can predict $\hat{y}$ as

$$\mathrm{G}_2 = (\lambda y)^{\gamma} \sim \mathrm{Gamma}\left(\frac{\alpha}{\gamma}, 1\right), \quad \hat{y}_{ij} = \frac{\mathrm{G}_2^{\frac{1}{\gamma}}}{\lambda_{ij}}.$$

After predicting the responses, the family of poverty measures for small area i, with poverty threshold of z is given by

$$P_{\alpha i} = \frac{1}{N_i}\sum_{j=1}^{N_i}\left(\frac{z-\hat{y}_{ij}}{z}\right)^{\alpha} I\left(E_{ij} < z\right), \qquad \alpha \geq 0, \quad i = 1, \cdots, A < L.$$

4.2.3 Record linkage of Sampling Units

We know the consumption values for all households surveyed in NLSS-II, therefore we do not need to predict them again. However, there is problem with the households *"record linkage"* between the NLSS-II and the population census data. We know that households in NLSS-II surveyed are also enumerated in the population census. We have information of their geographical location (wards in our case) both in survey and census, but we could not identify the same household in these two surveys. Since there is no record linkage between households, we consider all households in the census as non-sampled and predict responses for all. In our study, since the sample size is very small (see Table 1.1), it does not affect our estimations.

4.2.4 Small Area Estimation at District Level

We provide the SAE for poverty indicators at the district level for the *Mountains stratum* as an example. We have also estimated indicators in municipalities/VDC level and ward levels but are not tabulated here, there are large numbers of those small areas. We have mapped the poverty indicators at the district level. It would be better to provide maps of these indicators in the municipality/VDC level, unfortunately we do not have *shape file* in municipality/VDC level for mapping. As mentioned before we have shown an example with Mountains stratum in district level and all other stratum and SAE in different levels can be done similarly.

We have calculated the weighted estimates of the poverty indicators by direct method. Note, there are very small sample sizes for providing estimates in district level and two districts *"Rasuwa"* and *"Mustang"* are not in the sample. Because of very small sample sizes these estimates use to be misleading. We have used the Bayesian bootstrap sampling method with prior Dirichlet($\mathbf{0}$) for the direct estimates and are listed at the bottom of each table 4.2 , 4.3 , and 4.4.

We have also calculated the ELL method estimates using the nine covariates. In these nine covariates, four covariates *"skids714", "remtab", "hutype3"* and *"huown2"* are insignificant in the multivariate linear regression model at Mountasins stratum. In this dissertation, we have chosen nine covariates considering the whole nation and used them to develop models in all stratum. For the ELL estimates, we have used all nine covariates though there are insignificant covariates. For the heteroscadesticity model of the household error, we have two covariates "Urban and agri area 0.1+ Ha" *(nagar2)* and "proportion of own TV in ward" *(tvw)*. The household error variance modeling is done as it was carried out by Haslett et al. (2006). Note, SAE carried out in 2006 by CBS, the World Food Programme and the WB have used 37 covariates for SAE. We have listed the estimates of poverty indicators by ELL method at the end of each table 4.2, 4.3 and 4.4.

Table 4.2 provides Mountains stratum district level poverty incidence estimates for both selected models. The *noiseless model* has estimated the poverty incidence of the Mountain stratum at 0.424 with a standard error of 0.034. The *noisy model* has estimated the poverty incidence of the Mountain stratum at 0.364 with a standard error of 0.035. The *noisy model* has estimated less poverty incidence than the *noiseless model* and with similar standard errors. We take note that the *noiseless model* has estimated higher poverty incidences than the *noisy model* for mid-western and far-western regions. The Ell method has a smaller poverty rates and smaller standard errors than noiseless model for eastern, central and western districts of the Mountains stratum.

In the eastern districts (Taplejung, Sankhuwasabha and Solukhumbu) the *noiseless model* has estimated poverty incidences at 0.410 *(SE 0.034)*, 0.419 *(SE 0.036)* and 0.338 *(SE 0.031)*; and the *noisy model's* estimated poverty incidences of these districts are at 0.391 *(SE 0.036)*, 0.393 *(SE 0.036)* and 0.327 *(SE 0.036)* respectively. From the *noiseless model*, far-western districts (Bajura, Bajhang, and Darchula) have estimated poverty incidences at 0.457 *(SE 0.036)*, 0.474 *(SE 0.035)* and 0.491 *(SE 0.038)*; and *noisy model's* estimated poverty incidences to these districts are at 0.352 *(SE 0.035)*, 0.360 *(SE 0.037)* and 0.384 *(SE 0.038)* respectively. Figure 4.1 shows the map of the poverty incidences from the *noiseless model* and Figure 4.2 shows the map of the poverty incidences from the *noisy model*.

Table 4.3 provides the Mountains stratum district level poverty gap estimates for both selected models. The *noiseless model* has estimated the poverty gap of the Mountains stratum at 0.147 with a standard error of 0.015. The *noisy model* has estimated the poverty gap of the Mountain stratum at 0.102 with a standard error of 0.013. The *noisy model* has estimated poverty gap smaller. The Ell method has smaller poverty gaps and smaller standard errors than noiseless model for eastern, central and western districts of the Mountains stratum.

In the eastern districts (Taplejung, Sankhuwasabha and Solukhumbu) the *noiseless model* has estimated poverty gaps at 0.141 *(SE 0.015)*, 0.144 *(SE 0.016)* and 0.110 *(SE 0.013)*; and *noisy model's* estimated poverty gaps of these districts are at 0.112 *(SE 0.014)*, 0.112 *(SE 0.014)* and 0.087 *(SE 0.013)* respectively. From the *noiseless model*, far-western districts (Bajura, Bajhang, and Darchula) have estimated poverty gaps are at 0.162 *(SE 0.016)*, 0.171 *(SE 0.016)* and 0.178 *(SE 0.018)* and the *noisy model* has estimated poverty gaps to these districts are at 0.095 *(SE 0.013)*, 0.098 *(SE 0.014)* and 0.109 *(SE 0.015)* respectively. Figure 4.3 shows the map of the poverty gaps by the *noiseless model* and Figure 4.4 shows the map of the poverty gaps by the *noisy model*.

Table 4.4 provides Mountains stratum district level poverty severity estimates for both selected models. The *noiseless model* has estimated the poverty severity of the Mountains stratum at 0.072 with a standard error of 0.008; and, the *noisy model* has estimated the poverty severity of the Mountains stratum at 0.041 with a standard error of 0.006. Table 4.3 shows that poverty gaps estimated by the *noisy model* are smaller than for the *noiseless model*. Therefore, Table 4.4 also shows that poverty severities estimated by *noisy model* are smaller than *noiseless model*. The Ell method has smaller poverty severities than noiseless model estimates and their standard errors are generally smaller than noiseless model.

In the eastern districts (Taplejung, Sankhuwasabha and Solukhumbu) the *noiseless model* has estimated poverty severities at 0.068 *(SE 0.008)*, 0.070 *(SE 0.009)* and 0.052 *(SE 0.007)*. The *noisy model* has estimated poverty severities in these districts are at 0.045 *(SE 0.007)*, 0.045 *(SE 0.007)* and 0.033 *(SE 0.006)* respectively. From the *noiseless model*, far-western districts (Bajura, Bajhang, and Darchula) have estimated poverty severities 0.080 *(SE 0.009)*, 0.085 *(SE 0.009)* and 0.089 *(SE 0.010)*. From the *noisy model*, estimated poverty severities to these districts are at 0.037 *(SE 0.006)*, 0.038 *(SE 0.006)* and 0.044 *(SE 0.007)* respectively. Figure 4.5 shows the map of poverty severities from the *noiseless model*, and Figure 4.6 shows the map of poverty severities from the *noisy model*.

4.3 Simulation Study

We have covariates available in the census data, but we do not know their welfare consumption response values. So we generated response values for all households in the census data and applied the best fitted models we have developed, to see if the models can predict them. Now we will show the simulation study for Mountains stratum. We give the results of predicted poverty indicators for simulated census response data in district, municipality/VDC, and ward levels.

We simulate the response values in the census data with a multivariate linear regression using the nine covariates as we have used for the model building. They are (i) “Household size” (*hhsize*), (ii) “proportion of kids aged 0 - 6 in the household” (*skids6*), (iii) “proportion of kids aged 7 - 14 in the household” (*skids714*), (iv) “abroad migrant” (*remtab*), (v) “House temporary” (*hutype3*), (vi) “House owned” (*huown2*), (vii) “proportion of households with cooking fuel LP/gas in Ward" (*ckfuel3w*), (viii) "proportion of household with land-owning females in municipality/VDC" (*pflandv*), and (ix) "proportion of kids 6-16 attending school in municipality/VDC" (*pschv*). To simulate responses in the census, we fit the multivariate linear regression model with these nine covariates in NLSS-II data with log-transformed responses

$$\log(y_{ij}) = \boldsymbol{x'}_{ij}\boldsymbol{\beta} + e_{ij},\ i = 1, \cdots, \ell,\ j = 1, \cdots, n_i.$$

Then we predict simulated responses in the census census data

$$y_{ij}^{(s)} = e^{\boldsymbol{x'}_{ij}\hat{\boldsymbol{\beta}} + \hat{e}_{ij}},\ i = 1, \cdots, \mathrm{L},\ j = 1, \cdots, N_i,$$

where $\hat{e}_{ij}$ are generated from the assumption that residuals are distributed normally.

After generating simulated responses in the census data, we draw a samples of size n from the census data with simulated responses. We picked the same wards for the simulated sample as it was in NLSS-II data and the same number of households (12 households) by systematic random sampling as it was done in the NLSS-II. There are four Wards of NLSS-II (Mountains stratum), two wards from the Dolakha *(ward codes 2200606 and 2204407)* district and one ward from the Sankhuwasabha *(ward code 901303)* and Kalikot *(ward code 6400401)* districts, where the 2001 population census was unable to enumerate because of a Maoist insurgency at that time in those wards. For those wards where census data are not available, we randomly replace with the next ward from the same district. So in total we have 384 households from 32 wards in our simulated samples, as in NLSS-II.

We fit two models to the simulated sample data, the *noiseless model* and the *noisy model* with random area effects. Then we predict the responses in the census data from the models we fitted. If the models fit well, then they should predict the poverty indicators well.

We have calculated the three poverty indicators for both simulated responses and predicted responses using the poverty threshold line of 7,696 Nepalese rupees, the threshold as used for SAE Nepal (Haslett et al. 2006).

Figures 4.7 through 4.24 show diagonal plots of the poverty indicators from the simulation study in district, municipality/VDC and ward levels. The census poverty indicators are calculated from the simulated responses and the predicted poverty indicators are calculated from the two models we have fitted. We have also provided the linear relationship between census simulated indicators and model predicted indicators with their R^2 values in their respective plots.

Figures 4.7 and 4.8 contain simulation study poverty incidences for the district level from *noiseless* and *noisy* model respectively. The R^2 values of the linear relationship are 0.952 and 0.926 respectively. Figures 4.9 and 4.10 are simulation study poverty incidences for municipality/VDC level, from *noiseless* and *noisy* model respectively. The R^2 values of the linear relationship are 0.909 and 0.843 respectively. Figures 4.11 and 4.12 are simulation study poverty incidences at the ward level, from *noiseless* and *noisy* model respectively. The R^2 values of the linear relationship are 0.677 and 0.659 respectively.

Figures 4.13 through 4.18 show diagonal plots of the poverty gaps from the simulated responses versus predicted poverty gaps from two fitted models in district, municipality/VDC and ward levels. We have also provided the linear relationship between the simulated response gaps in the census data and predicted responses gap by the fitted model with the R^2 values in their respective plots.

Figures 4.13 and 4.14 are simulation study poverty gaps at the district level from *noiseless* and *noisy* model respectively. Figures 4.15 and 4.16 are simulation study poverty gaps in municipality/VDC level; figures 4.17 and 4.18 are simulation study poverty gaps in the ward level from *noiseless* and *noisy* model respectively.

Figures 4.19 through 4.24 show diagonal plots of the poverty severities from the simulated responses in the census data versus predicted poverty severities from two fitted models in district, municipality/VDC and ward levels. We have also provided the linear relationship between the simulated poverty severities in the census data and predicted poverty severities from the fitted model with their respective R^2 values.

Figures 4.19 and 4.20 are a simulation study of poverty severities at the district level from *noiseless* and *noisy* model respectively. In Figure 4.19 all the observations are above the diagonal plot showing that the generalized gamma model predicts larger severities than the true value while in Figure 4.20 there are some observations below the diagonal line, though it is also up-lifted. Examining these two plots of the simulation study on severities, it shows that the mixture of two generalized gamma (GB2) model is better than the generalized gamma model. Figure 4.21 and 4.22 are simulation studies on poverty severities at the municipality/VDC level from *noiseless* and *noisy* model respectively; Figures 4.23 and 4.24 are simulation study poverty severities in the ward level from *noiseless* and *noisy* model respectively.

4.4 Conclusion

Now, we summarize our research contributions in modeling continuous and positively skewed (CPS) size data without the logarithmic transformation and discuss its applications and future work.

4.4.1 Contributions in Methodology

The goal of this dissertation is to fit hierarchical Bayesian models without logarithmic transformation to CPS data from small areas and to introduce covariates into the model. We build models in the survey data and link the survey and census data for the prediction of responses in the census. We have chosen the positively skewed density functions for modeling with non-informative priors, except for one shape parameter of the generalized gamma distribution. The best fitted models are chosen using LPML criterion. Our target is to apply the best fitted model for SAE. We have demonstrated our application to welfare consumption data from NLSS-II and the population census data. These models can be applied to any other CPS size data.

Prior to model building, first we need the assumption of responses data as noiseless or noisy. Our idea is to select a density function for the model with the possibility of noise being introduced in the responses or not. We have chosen the generalized gamma distribution and its special cases for noiseless data modeling. Let the response variable have the generalized gamma distribution

$$\begin{aligned} y|\alpha,\lambda,\gamma &\overset{ind}{\sim} \text{GGamma}\,(\alpha,\lambda,\gamma)\,, \quad \lambda,\alpha,\gamma>0,\ y>0, \\ f(y|\alpha,\lambda,\gamma) &= \gamma\,\frac{e^{-\lambda y^{\gamma}}\ y^{\alpha\gamma-1}}{\Gamma(\alpha)}\,\lambda^{\alpha}. \end{aligned} \tag{4.1}$$

The k^{th} moment of the response variable is given by

$$E[Y^k|\alpha,\lambda,\gamma] = \frac{\Gamma\left(\frac{\alpha+k}{\gamma}\right)}{\Gamma\left(\frac{\alpha}{\gamma}\right)}\,\lambda^{-k}. \tag{4.2}$$

If the responses are noisy, then we have chosen the GB2 distribution and its special cases for modeling. The intuition behind using the GB2 for noisy responses is that, in the GB2 distribution the true rate parameter of the response variable is hidden. This rate parameter has the next generalized gamma distribution. So our belief is that it is useful for modeling noisy data. Let, the response variable Y have the generalized gamma distribution with rate λ and this rate parameter have the next generalized gamma distribution with rate θ

$$y|\lambda,\alpha,\gamma \overset{ind}{\sim} \text{GGamma}(\alpha,\lambda,\gamma), \quad \lambda|\theta,\phi,\gamma \overset{iid}{\sim} \text{GGamma}(\phi,\theta,\gamma).$$

Mixing these two generalized gamma distributions yields a GB2 distribution

$$y|\alpha,\phi,\theta,\gamma \overset{ind}{\sim} \text{GB2}(\alpha,\phi,\theta,\gamma).$$

Its k^{th} moment,

$$E[Y^k|\alpha,\phi,\theta,\gamma] = \frac{\Gamma\left(\frac{\alpha+k}{\gamma}\right)}{\Gamma\left(\frac{\alpha}{\gamma}\right)}\frac{\Gamma\left(\frac{\phi-k}{\gamma}\right)}{\Gamma\left(\frac{\phi}{\gamma}\right)}\,\theta^{k}, \quad \phi>k, \tag{4.3}$$

depends on the parameters of the distribution of the rate parameter λ.

After fitting the models, we provided the small area estimates. To facilitate the prediction of the responses given covariates for the SAE, we have introduced the covariates in the models through their rate parameters. We built the models both with and without random area effects. If random area effects are not important, then we can have models without random area effects. The models without random area effects are much simpler. In our application problem, we found that random area effect parameters improve our models much more and therefore are essential.

We fitted the generalized gamma distribution and its special cases to model the noiseless responses starting from a simple to a more generalized distribution: the exponential, the gamma and the generalized gamma. In our application, we noticed that the model with a more generalized distribution fits better. The generalized gamma model is the best fit for our application with a competitive gamma model. Assuming noisy responses, we have fitted the GB2 distribution and its special cases: the mixture of exponential and gamma distributions, the mixture of two gamma distributions and the mixture of two generalized gamma distributions. In our application, as in the noiseless responses, the mixture of two generalized gamma GB2 model is the best fit. However, the mixture of two gamma GB2 model is competitive. Regarding random area effects, in our application, whether responses are assumed noisy or not, we found that all models with random area effects are better than those without random area effects.

The Metropolis–Hastings algorithm is the main sampling procedure we have used for sampling parameters. We have an approximated multivariate normal distribution for $\boldsymbol{\beta}$ and $\boldsymbol{\nu}$, but we do not know the distribution of the shape and rate parameters. We approximated their joint distribution by the multivariate normal distribution according to their logarithmic-transformed variable. In our computation, we used a clever way of getting the approximated mean vector and the covariance matrix from the available samples in the previous step since our belief is that it is the closest one for the parameter to approximate. If there is no chance to draw samples in an easier and faster way, then we draw samples using the grid method.

4.4.2 Application

The hierarchical Bayesian models we have developed can be applied for modeling any CPS size data. This model could also be helpful for any data with a heavy-tailed distribution since the generalized gamma or GB2 are good for working heavy-tailed distributions, such as insurance data or income data.

We have shown one application in welfare consumption data from NLSS-II to estimate poverty indicators in the small areas. For SAE we link the survey and the census data. We carry the model parameters from the survey to the census data. In our application models with gamma and generalized gamma distributions, both fit closely and competitively, whether assuming noisy or noiseless responses. However, the generalized gamma model has been shown to be a little better. We have also shown the simulation results in Chapter 4 for modeling noiseless data by using the generalized gamma distribution with random area effects and for noisy data by using the mixture of two generalized gamma GB2 distributions with random area effects. The simulation study supports our study that we can get better estimates of the responses without log transformation. In our simulation study we have a higher linear relationship between the simulated response and the predicted response's R^2 value for poverty incidence in the district level, then next higher R^2 value for the municipality/VDC level, then the ward level. Looking at the figures for district level poverty severities in simulation data, it shows that assumption of noisy responses fits well.

4.4.3 Future work

Our study can be extended to cover many new situations. Some of the possible future works are as follows.

Spatial analysis

In our models we found that random area effects are important parameters, and we can add spatial effects as mentioned by He and Sun (2000). We will summarize the idea from their paper. To accommodate spatial effects, we can model with the conditional auto regressive (CAR) model of Clayton and Kaldor (1987). Let $\nu_i,\ i = 1 \cdots, \ell$, be the random area effects and its the adjacency matrix (C_{kl}), the adjacency matrix is a (0,1)-matrix with 1 for adjacent area, 0 for non-adjacent area and zeros in its diagonal. The symmetric adjacency matrix $C = (C_{kl})$, with eigenvalues $\lambda_1, \leq \cdots \leq \lambda_\ell,\ \lambda_1 < 0, \lambda_\ell > 0$, follow the normal distribution

$$\nu_i | \nu_k, k \neq i \sim N\left(\rho \sum_{k \neq i} C_{ik} \nu_k,\ \delta\right).$$

If $\lambda_1^{-1} \le \rho \le \lambda_\ell^{-1}$, then matrix B given below is positive definite (Besag, 1974)

$$B = I - \rho C,$$

where I is $\ell \times \ell$ identity matrix, and joint distribution of the random area effect $\boldsymbol{\nu}$ is given by the multivariate normal distribution

$$\boldsymbol{\nu} \sim N\left(\mathbf{0},\ \delta B^{-1}\right).$$

When B is positive semi-definite, the joint density of $\boldsymbol{\nu}$ is called the partially informative normal distribution. In our model rather than assuming $\nu_i \overset{\text{iid}}{\sim} N(0, \sigma^2)$, $i = 1 \cdots, \ell$, we can use the CAR model with spatial correlation to improve the model.

Robustness of area effects

For the robustness of the model, we can place a Dirichlet process (DP) on the random area effects as

$$\begin{aligned}
&\nu_i|\psi_i \overset{\text{ind}}{\sim} N(\psi_i,\, (1-\kappa)\delta^2), \quad i = 1, \cdots, \ell, \\
&\psi_i|G \overset{\text{ind}}{\sim} G, \\
&\quad G \sim \text{DP}\,(\alpha,\, G_0)\,,
\end{aligned}$$

where α is a concentration parameter with the baseline distribution $G_0 = \Phi\left(\frac{t}{\sqrt{\kappa}\,\delta}\right)$ and κ is specified. This model has one more level, distribution of ψ_i, which allow us to jitter the random area effects ν_i in the Dirichlet process.

4.4.4 Nested multi-level effects

In this dissertation, we have the area effects in the PSU level. However we could introduce the multi-level area effects. Suppose, we have first level area effects τ_i, second level area effects ω_{ij} and third level area effects $\nu_{ijk}, i = 1, \cdots, I,\ j = 1, \cdots, J_i,\ k = 1, \cdots, K_{ij}$. We assume that these first, second and third level area effects have a normal distribution with mean zero and constant variances σ_I^2, σ_J^2, and σ_K^2 respectively. If we have noisy responses then the multi-level model can be written as

$$\begin{aligned}
y_{ijkl}|\boldsymbol{\beta},\alpha,\gamma,\tau_i,\omega_{ij},\nu_{ijk} &\overset{\text{ind}}{\sim} \text{GB2}\left(\alpha,\ e^{\boldsymbol{x}'_{ijkl}\boldsymbol{\beta}+\boldsymbol{\tau_i}+\boldsymbol{\omega_{ij}}+\boldsymbol{\nu_{ijk}}},\ \gamma\right),\\
\pi(\boldsymbol{\beta},\alpha,\sigma_I^2,\sigma_J^2,\sigma_K^2) &\propto \frac{1}{(1+\alpha)^2\,(1+\sigma_I^2)^2\,(1+\sigma_J^2)^2\,(1+\sigma_K^2)^2}\\
\gamma &\sim \text{Gamma}(S,R), \quad \text{where 'S' and 'R' are specified.}
\end{aligned}$$

We can write a similar model for noiseless responses also. This model is more complicated since we have more parameters in different levels and posterior distribution and conditional posterior distribution are not in simple form.

Different gamma parameters for each area

In our GB2 models, we can create a more pronounced area effect by considering each area as having a different gamma parameter γ_i, $i=1,\cdots,\ell$. Then our GB2 likelihood will be

$$\pi(\boldsymbol{y}|\alpha,\boldsymbol{\beta},\boldsymbol{\nu},\boldsymbol{\gamma}) = \prod_{i=1}^{\ell}\prod_{j=1}^{n_i}\gamma_i\frac{y_{ij}^{\alpha-1}}{\mathrm{B}(\frac{\alpha}{\gamma_i},\frac{\alpha+2}{\gamma_i})}\frac{e^{-\alpha(\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\nu_i)}}{\left(1+\left[y_{ij}\,e^{-(\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\nu_i)}\right]^{\gamma_i}\right)^{\frac{2(\alpha+1)}{\gamma_i}}}.$$

The mixture of two generalized gamma distributions (GB2) model is

$$\begin{aligned}
y_{ij}|\boldsymbol{\beta},\alpha,\gamma_i,\nu_i &\overset{\text{ind}}{\sim} \text{GB2}\left(\alpha, e^{\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\boldsymbol{\nu_i}},\gamma_i\right), \quad \theta_{ij}=e^{\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\boldsymbol{\nu_i}},\ i=1,\cdots,\ell,\ j=1,\cdots,n_i,\\
\nu_i &\overset{\text{iid}}{\sim} N(0,\sigma^2),\\
\pi(\boldsymbol{\beta},\alpha,\sigma^2) &\propto \frac{1}{(1+\alpha)^2\,(1+\sigma^2)^2},\\
\gamma_i &\sim \text{Gamma}(S_i,R_i), \quad \text{where shape } S_i \text{ and rate } R_i \text{ are specified.}
\end{aligned}$$

Here we use one γ_i parameter for each area, which may help to improve the model, although it demands more computation.

Mixture of two Gamma Distribution
With Different Shape Gamma Parameter

Up to now for the GB2 distribution, we have considered that the mixture of the distributions of responses $y|\alpha,\lambda,\gamma \sim \text{GGamma}(\alpha,\lambda,\gamma)$ and its rate parameter $\lambda|\theta,\phi,\gamma \sim$

GGamma(ϕ, θ, γ) both having the generalized gamma distribution, where both these distributions have the same shape parameter γ. A possibly interesting extension is when shape parameters are different.

Let us consider the mixture of two generalized gamma distributions as with different γ and ξ parameters as

$$y|\alpha, \lambda, \gamma \sim \text{GGamma}\left(\alpha, \lambda, \gamma\right) \quad \text{and} \quad \lambda|\theta, \phi, \xi \sim \text{GGamma}(\phi, \theta, \xi)$$

Mixing these two distributions and integrating out the rate parameter λ, we have

$$\begin{aligned} f(y|\alpha, \phi, \theta) &= \int_0^\infty f(y|\lambda, \alpha, \gamma) \;\; g(\lambda|\theta, \phi, \xi) \; d\lambda, \quad \alpha, \phi, \gamma, \xi, \theta > 0, \\ &= \frac{\gamma \xi y^{\alpha-1} \theta^{\phi}}{\Gamma\left(\frac{\alpha}{\gamma}\right) \Gamma\left(\frac{\phi}{\xi}\right)} \int_0^\infty e^{-(y^\gamma \lambda^\gamma + \theta^\xi \lambda^\xi)} \; \lambda^{\alpha+\phi-1} \, d\lambda. \end{aligned} \tag{4.4}$$

This density function is not easily integrable, however this is a GB2 distribution when $\gamma = \xi$. This density is a more general type of distribution than GB2 with one more parameter added. However it is not derived as a special case of Beta-F distribution as we have discussed in subsection *"Modeling with GB2 Distribution"* section *"Distributions Used in Model Building"*, Chapter 1. Clearly, this density is much more complicated.

Table 4.2: Poverty incidences by districts, SAE 2001

District	Region	Noiseless model				Noisy model			
		Rate	**SE**	**Hpd lower**	**Hpd Upper**	**Rate**	**SE**	**Hpd lower**	**Hpd Upper**
Taplejung	Eastern	0.410	0.034	0.343	0.475	0.391	0.036	0.324	0.463
Sankhuwasabha	Eastern	0.419	0.036	0.353	0.495	0.393	0.036	0.326	0.464
Solukhumbu	Eastern	0.338	0.031	0.276	0.397	0.327	0.036	0.257	0.395
Dolakha	Central	0.388	0.034	0.326	0.457	0.351	0.034	0.284	0.418
Sindhupalchok	Central	0.407	0.034	0.337	0.470	0.359	0.035	0.295	0.426
Rasuwa	Central	0.430	0.037	0.358	0.501	0.387	0.036	0.322	0.458
Manang	Western	0.431	0.038	0.350	0.496	0.379	0.038	0.313	0.456
Mustang	Western	0.450	0.037	0.375	0.520	0.411	0.036	0.348	0.489
Dolpa	Mid Western	0.411	0.035	0.336	0.476	0.335	0.035	0.264	0.399
Jumla	Mid Western	0.450	0.036	0.381	0.517	0.352	0.038	0.282	0.427
Kalikot	Mid Western	0.472	0.038	0.396	0.550	0.376	0.037	0.294	0.439
Mugu	Mid Western	0.439	0.035	0.369	0.510	0.357	0.035	0.292	0.428
Humla	Mid Western	0.457	0.035	0.386	0.520	0.351	0.036	0.285	0.422
Bajura	Far Western	0.457	0.036	0.391	0.530	0.352	0.035	0.285	0.417
Bajhang	Far Western	0.474	0.035	0.407	0.542	0.360	0.037	0.295	0.436
Darchula	Far Western	0.491	0.038	0.414	0.566	0.384	0.038	0.308	0.458
Mountains		**0.424**	**0.034**	**0.358**	**0.487**	**0.364**	**0.035**	**0.291**	**0.430**

The direct estimate poverty rates *(standard errors)* are: Taplejung 0.286 *(0.075)*, Sankhuwasabha 0.288 *(0.066)*, Solukhumbu 0.165 *(0.073)*, Dolakha 0.191 *(0.056)*, Sindhupalchok 0.432 *(0.053)*, Rasuwa — — Manang 0.000 *(0.000)*, Mustang — — Dolpa 0.000 *(0.000)*, Jumla 0.253 *(0.119)*, Kalikot 0.594 *(0.139)*, Mugu 0.298 *(0.124)*, Humla 0.912 *(0.076)*, Bajura 0.286 *(0.087)*, Bajhang 0.114 *(0.059)*, Darchula 0.535 *(0.102)*, and the Mountains stratum 0.326 *(0.023)*.
The ELL method poverty rate estimates *(standard errors)* are: Taplejung 0.286 *(0.0251)*, Sankhuwasabha 0.302 *(0.027)*, Solukhumbu 0.266 *(0.027)*, Dolakha 0.295 *(0.019)*, Sindhupalchok 0.340 *(0.020)*, Rasuwa 0.367 *(0.0237)*, Manang 0.261 *(0.041)*, Mustang 0.282 *(0.036)*, Dolpa 0.408 *(0.026)*, Jumla 0.505 *(0.030)*, Kalikot 0.505 *(0.035)*, Mugu 0.493 *(0.034)*, Humla 0.510 *(0.034)*, Bajura 0.462 *(0.030)*, Bajhang 0.503 *(0.0309)*, Darchula 0.401 *(0.040)*, and the Mountains stratum 0.369 *(0.021)* .

Table 4.3: Poverty gaps by districts, SAE 2001

District	Region	Noiseless model				Noisy model			
		Gap	**SE**	**Hpd lower**	**Hpd Upper**	**Gap**	**SE**	**Hpd lower**	**Hpd Upper**
Taplejung	Eastern	0.141	0.015	0.113	0.170	0.112	0.014	0.083	0.137
Sankhuwasabha	Eastern	0.144	0.016	0.115	0.178	0.112	0.014	0.083	0.137
Solukhumbu	Eastern	0.110	0.013	0.087	0.134	0.086	0.013	0.061	0.110
Dolakha	Central	0.130	0.015	0.105	0.159	0.097	0.013	0.072	0.121
Sindhupalchok	Central	0.139	0.015	0.108	0.166	0.101	0.013	0.078	0.127
Rasuwa	Central	0.149	0.017	0.114	0.178	0.117	0.014	0.092	0.148
Manang	Western	0.153	0.017	0.120	0.185	0.106	0.014	0.082	0.138
Mustang	Western	0.170	0.018	0.137	0.205	0.132	0.014	0.104	0.160
Dolpa	Mid Western	0.143	0.015	0.110	0.169	0.090	0.013	0.067	0.117
Jumla	Mid Western	0.159	0.016	0.130	0.189	0.094	0.014	0.069	0.123
Kalikot	Mid Western	0.168	0.018	0.131	0.200	0.103	0.014	0.075	0.130
Mugu	Mid Western	0.154	0.016	0.126	0.187	0.098	0.013	0.074	0.123
Humla	Mid Western	0.163	0.016	0.134	0.194	0.095	0.013	0.073	0.122
Bajura	Far Western	0.162	0.016	0.131	0.193	0.095	0.013	0.071	0.120
Bajhang	Far Western	0.171	0.016	0.141	0.202	0.098	0.014	0.075	0.127
Darchula	Far Western	0.178	0.018	0.146	0.214	0.109	0.015	0.081	0.137
Mountains		**0.147**	**0.015**	**0.121**	**0.177**	**0.102**	**0.013**	**0.075**	**0.128**

The direct estimate poverty gap estimates *(standard errors)* are: Taplejung 0.0272 *(0.0077)*, Sankhuwasabha 0.0846 *(0.0237)*, Solukhumbu 0.0270 *(0.0135)*, Dolakha 0.0283 *(0.0101)*, Sindhupalchok 0.0967 *(0.0168)*, Rasuwa — — Manang 0.000 *(0.000)*, Mustang — — Dolpa 0.000 *(0.000)*, Jumla 0.0422 *(0.0257)*, Kalikot 0.1221 *(0.0412)*, Mugu 0.0531 *(0.0228)*, Humla 0.2639 *(0.0411)*, Bajura 0.0772 *(0.0280)*, Bajhang 0.0330 *(0.0241)*, Darchula 0.0869 *(0.0262)*, and the Mountains stratum 0.0695 *(0.0066)*. The ELL method poverty gap estimates *(standard errors)* are: Taplejung 0.0656 *(0.0084)*, Sankhuwasabha 0.0829 *(0.0101)*, Solukhumbu 0.0591 *(0.0080)*, Dolakha 0.0757 *(0.0074)*, Sindhupalchok 0.0868 *(0.0088)*, Rasuwa 0.0958 *(0.0104)*, Manang 0.0686 *(0.0155)*, Mustang 0.0864 *(0.0137)*, Dolpa 0.1062 *(0.0112)*, Jumla 0.1456 *(0.0155)*, Kalikot 0.1406 *(0.0165)*, Mugu 0.1376 *(0.0163)*, Humla 0.1435 *(0.0168)*, Bajura 0.1222 *(0.0136)*, Bajhang 0.1413 *(0.0156)*, Darchula 0.1028 *(0.0156)*, and the Mountains stratum 0.0971 *(0.009)*.

Table 4.4: Poverty severities by districts, SAE 2001

District	Region	Noiseless model				Noisy model			
		Severity	**SE**	**Hpd lower**	**Hpd Upper**	**Severity**	**SE**	**Hpd lower**	**Hpd Upper**
Taplejung	Eastern	0.068	0.008	0.052	0.084	0.045	0.007	0.032	0.058
Sankhuwasabha	Eastern	0.070	0.009	0.053	0.088	0.045	0.007	0.032	0.058
Solukhumbu	Eastern	0.052	0.007	0.038	0.065	0.033	0.006	0.022	0.044
Dolakha	Central	0.062	0.008	0.048	0.078	0.039	0.006	0.028	0.051
Sindhupalchok	Central	0.067	0.008	0.050	0.082	0.041	0.006	0.030	0.053
Rasuwa	Central	0.073	0.010	0.053	0.090	0.051	0.007	0.038	0.065
Manang	Western	0.076	0.010	0.057	0.095	0.042	0.007	0.030	0.057
Mustang	Western	0.092	0.011	0.072	0.113	0.063	0.007	0.049	0.078
Dolpa	Mid Western	0.070	0.008	0.055	0.088	0.036	0.006	0.024	0.048
Jumla	Mid Western	0.078	0.009	0.061	0.095	0.037	0.006	0.024	0.049
Kalikot	Mid Western	0.083	0.010	0.063	0.103	0.041	0.007	0.028	0.054
Mugu	Mid Western	0.076	0.009	0.060	0.095	0.039	0.006	0.027	0.051
Humla	Mid Western	0.081	0.009	0.066	0.100	0.037	0.006	0.026	0.049
Bajura	Far Western	0.080	0.009	0.064	0.098	0.037	0.006	0.026	0.049
Bajhang	Far Western	0.085	0.009	0.070	0.105	0.038	0.006	0.025	0.050
Darchula	Far Western	0.089	0.010	0.071	0.110	0.044	0.007	0.030	0.057
Mountains		**0.072**	**0.008**	**0.056**	**0.087**	**0.041**	**0.006**	**0.029**	**0.054**

The direct estimate poverty severities with *(standard errors)* are: Taplejung 0.0032 *(0.0012)*, Sankhuwasabha 0.0325 *(0.0099)*, Solukhumbu 0.0057 *(0.0033)*, Dolakha 0.0051 *(0.0025)*, Sindhupalchok 0.0336 *(0.0077)*, Rasuwa — — Manang 0.000 *(0.000)*, Mustang — — Dolpa 0.000 *(0.000)*, Jumla 0.0104 *(0.0072)*, Kalikot 0.0364 *(0.0216)*, Mugu 0.0092 *(0.0040)*, Humla 0.0911 *(0.0211)*, Bajura 0.0242 *(0.0106)*, Bajhang 0.0164 *(0.0135)*, Darchula 0.0250 *(0.0100)*, and the Mountains stratum 0.0221 *(0.0028)*. The ELL method poverty severity estimates with *(standard errors)* are: Taplejung 0.0228 *(0.0037)*, Sankhuwasabha 0.0364 *(0.0055)*, Solukhumbu 0.0203 *(0.0033)*, Dolakha 0.0308 *(0.0041)*, Sindhupalchok 0.0332 *(0.0046)*, Rasuwa 0.0360 *(0.0052)*, Manang 0.0277 *(0.0079)*, Mustang 0.0433 *(0.0078)*, Dolpa 0.0407 *(0.0057)*, Jumla 0.0595 *(0.0086)*, Kalikot 0.0550 *(0.0087)*, Mugu 0.0538 *(0.0086)*, Humla 0.0570 *(0.0091)*, Bajura 0.0468 *(0.0070)*, Bajhang 0.0560 *(0.0086)*, Darchula 0.0387 *(0.0075)*, and the Mountains stratum 0.0380 *(0.0050)*.

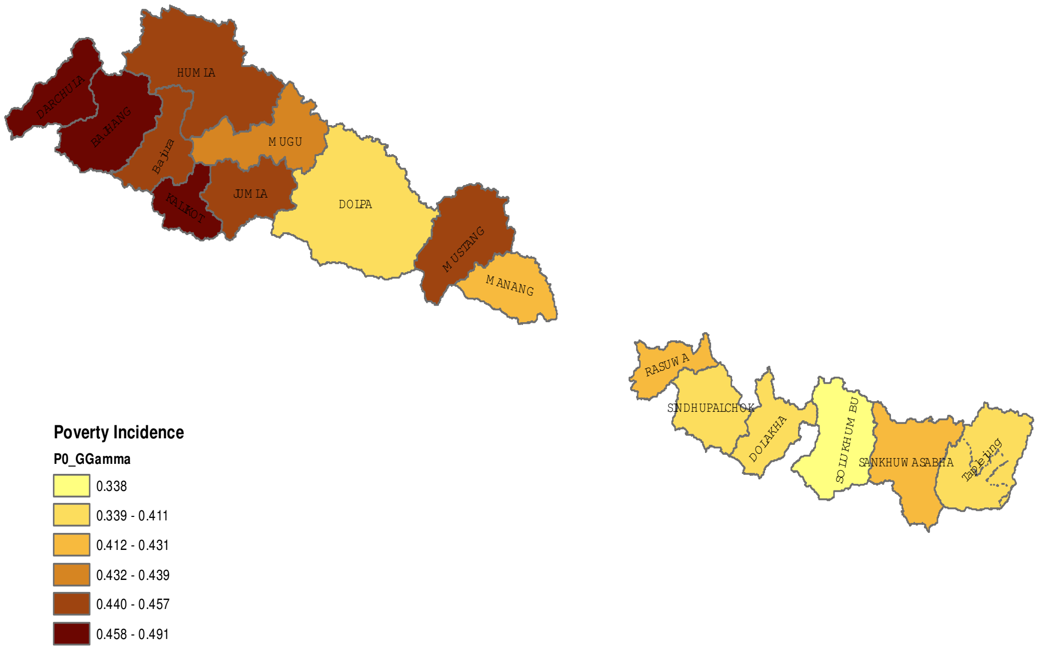


Figure 4.1: Poverty incidence (P0) at the district level (Noiseless model, Mountains stratum)

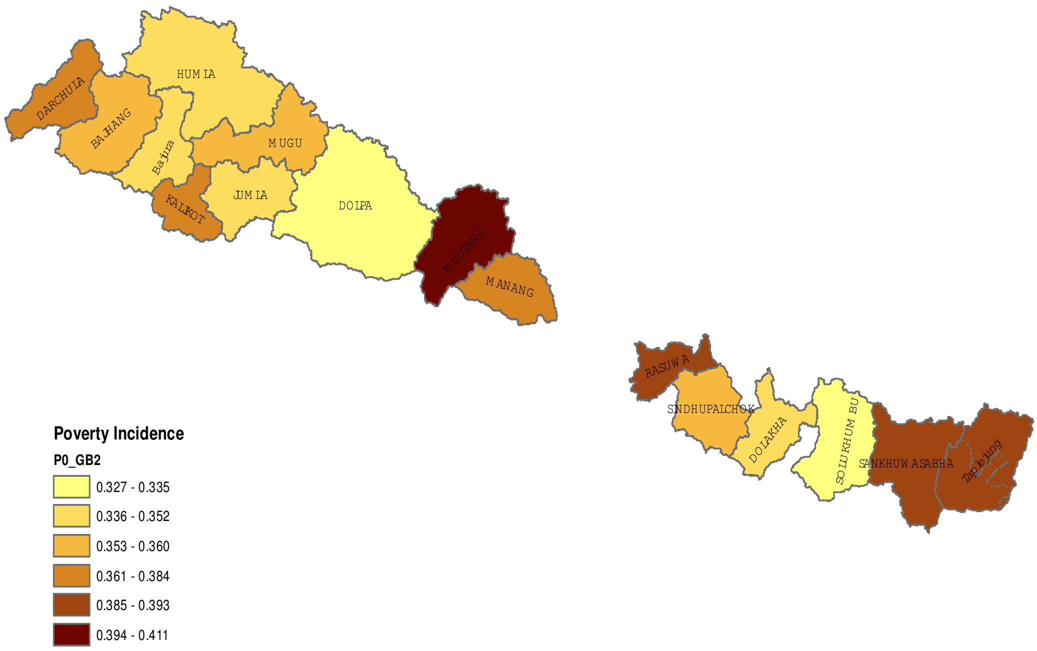


Figure 4.2: Poverty incidence (P0) at the district level (Noisy model, Mountains stratum)
Note: mapping categories are different than above figure.

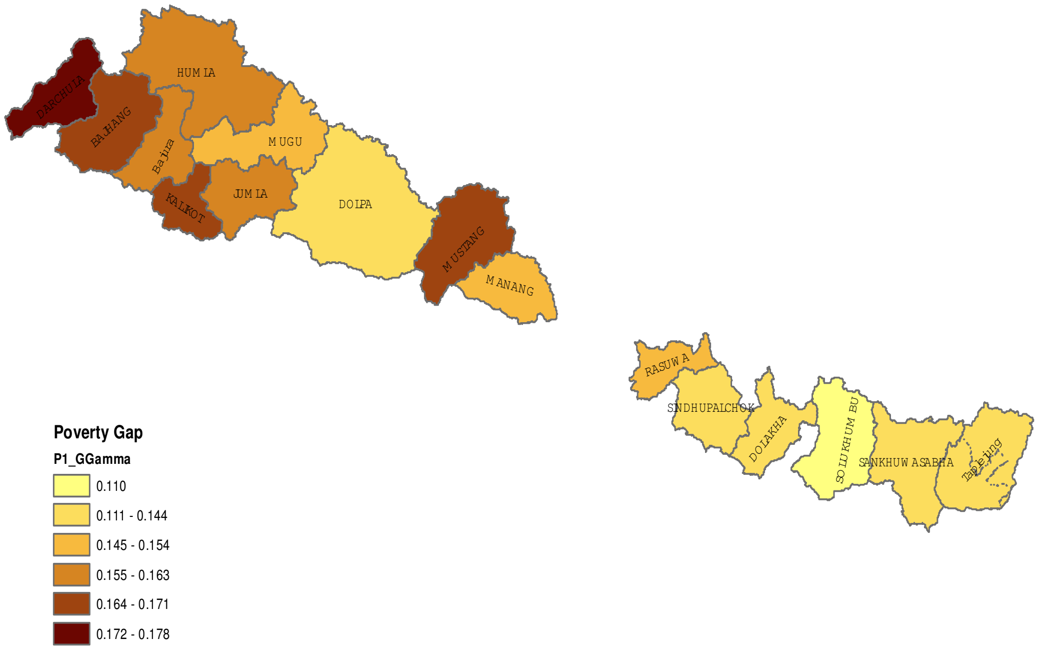


Figure 4.3: Poverty gap (P1) at the district level (Noiseless model, Mountains stratum)

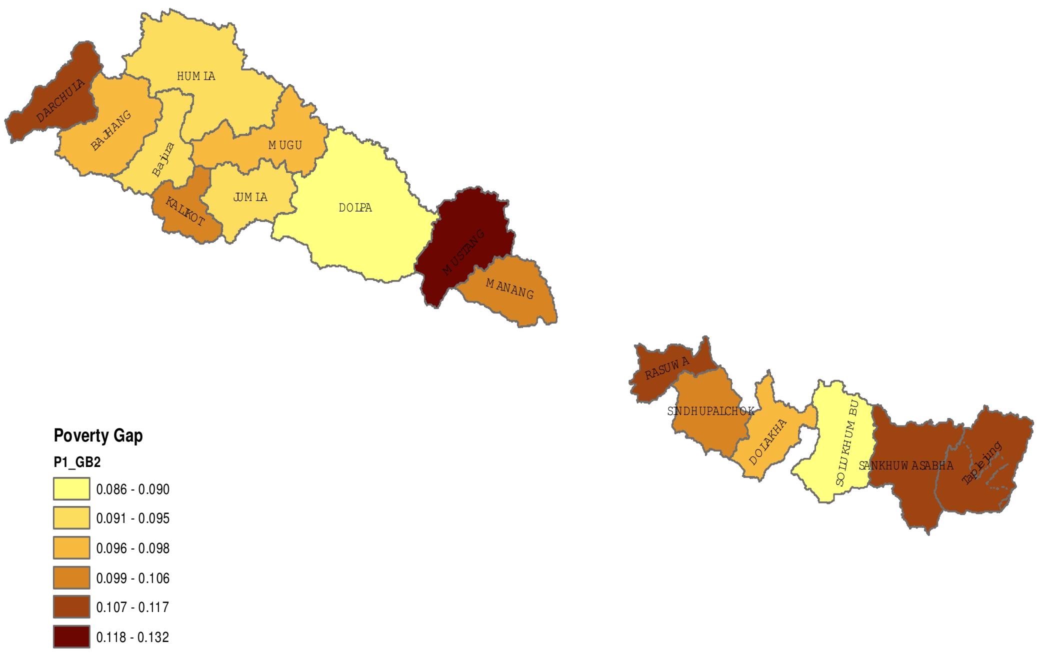


Figure 4.4: Poverty gap (P1) at the district level (Noisy model, Mountains stratum)
Note: mapping categories are different than above figure.

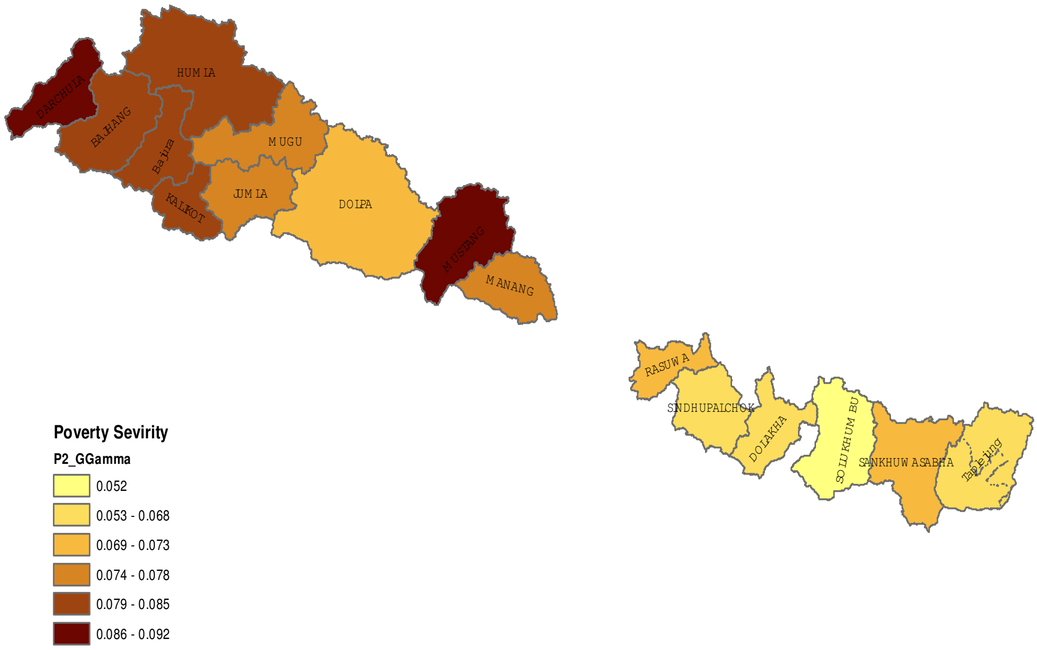


Figure 4.5: Poverty severity (P2) at the district level (Noiseless model, Mountains stratum)

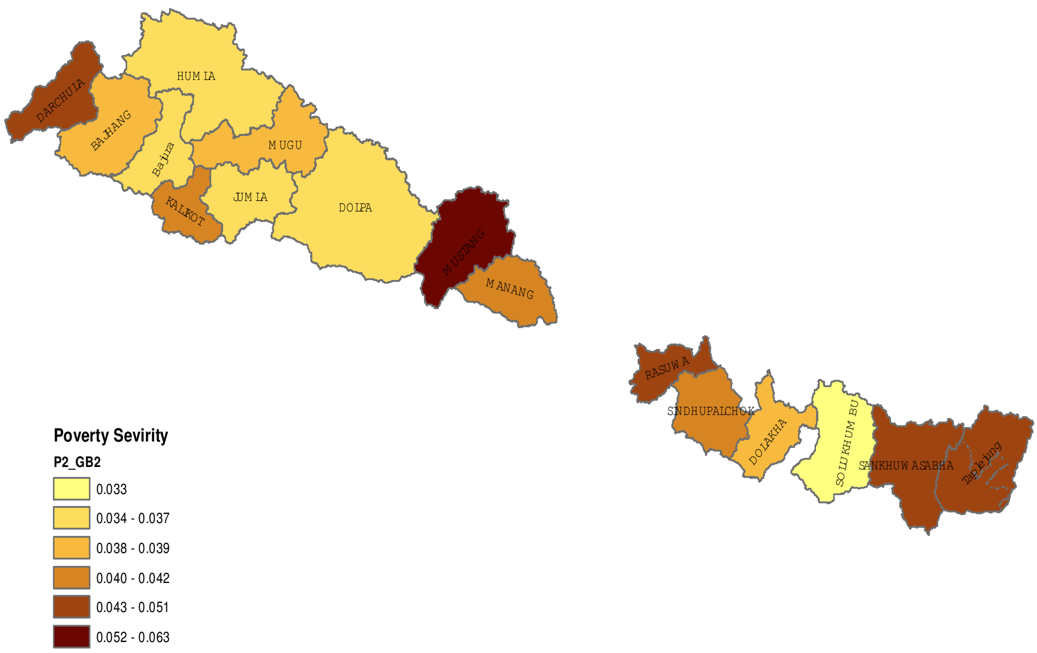


Figure 4.6: Poverty severity (P2) at the district level (Noisy model, Mountains stratum)
Note: mapping categories are different than above figure.

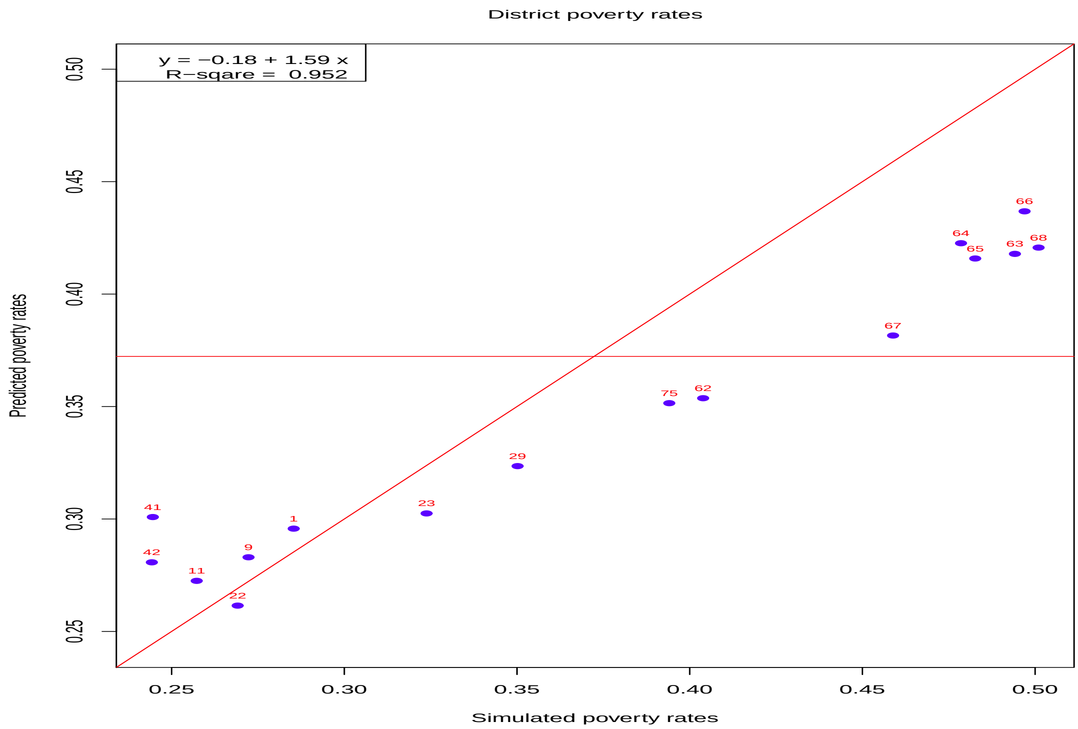


Figure 4.7: Poverty incidences in the simulation study by district (Noiseless model, Mountains stratum)

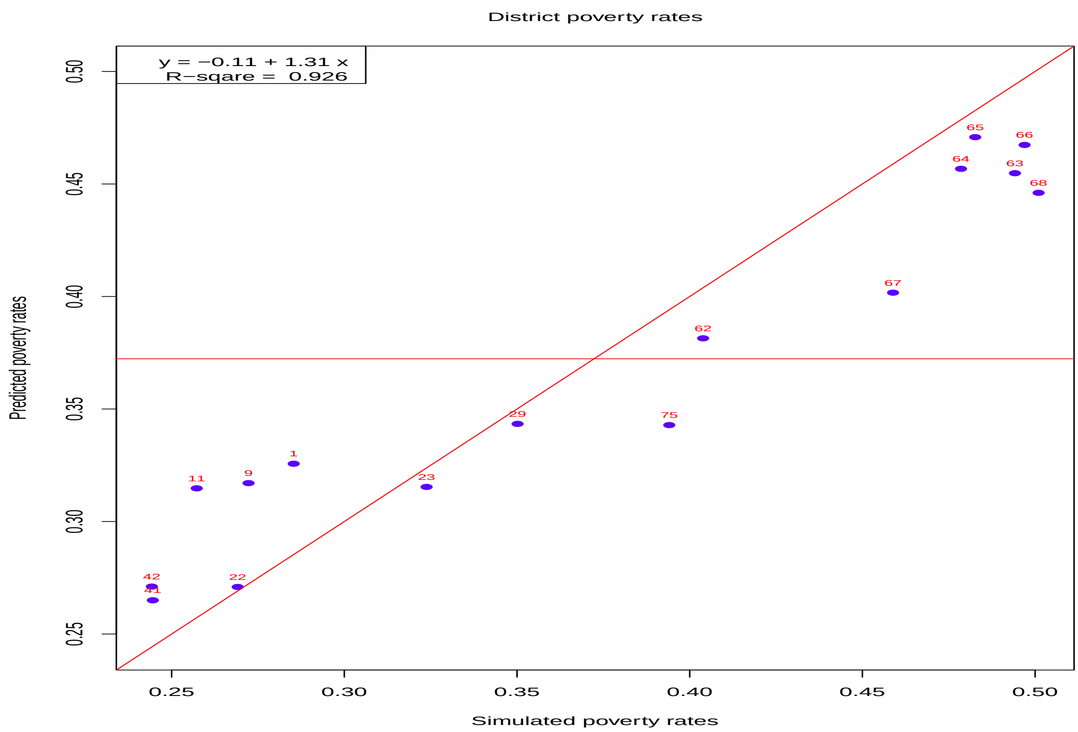


Figure 4.8: Poverty incidences in the simulation study by district (Noisy model, Mountains stratum)

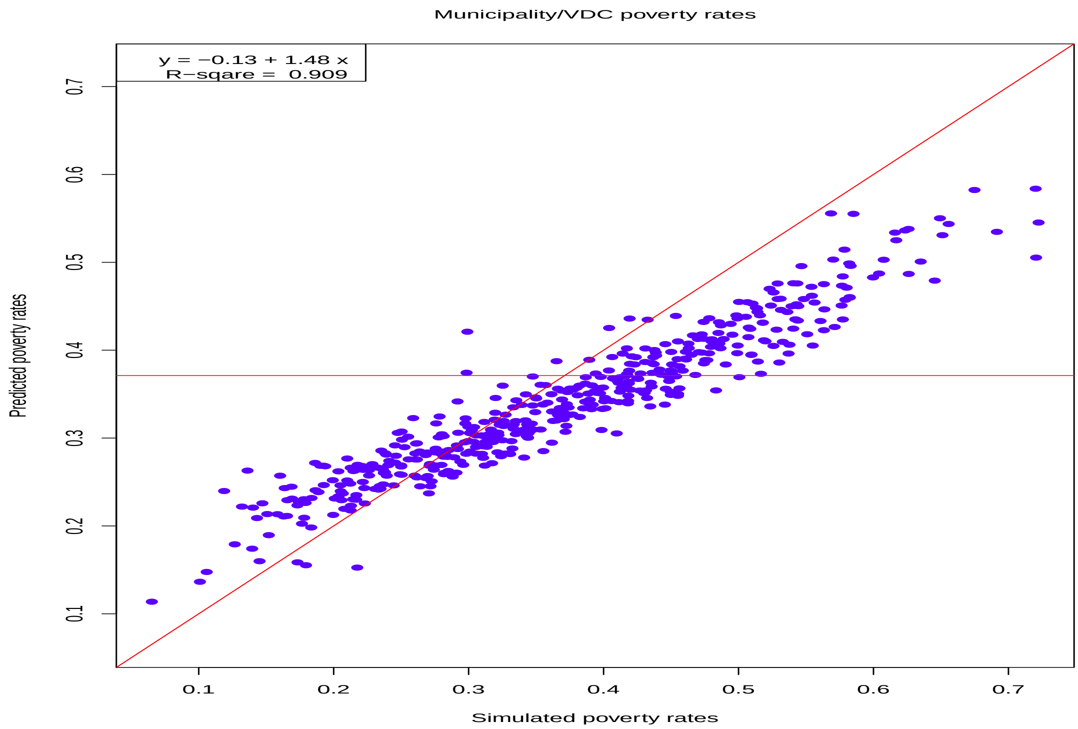


Figure 4.9: Poverty incidences in the simulation study by municipality/VDC (Noiseless model, Mountains stratum)

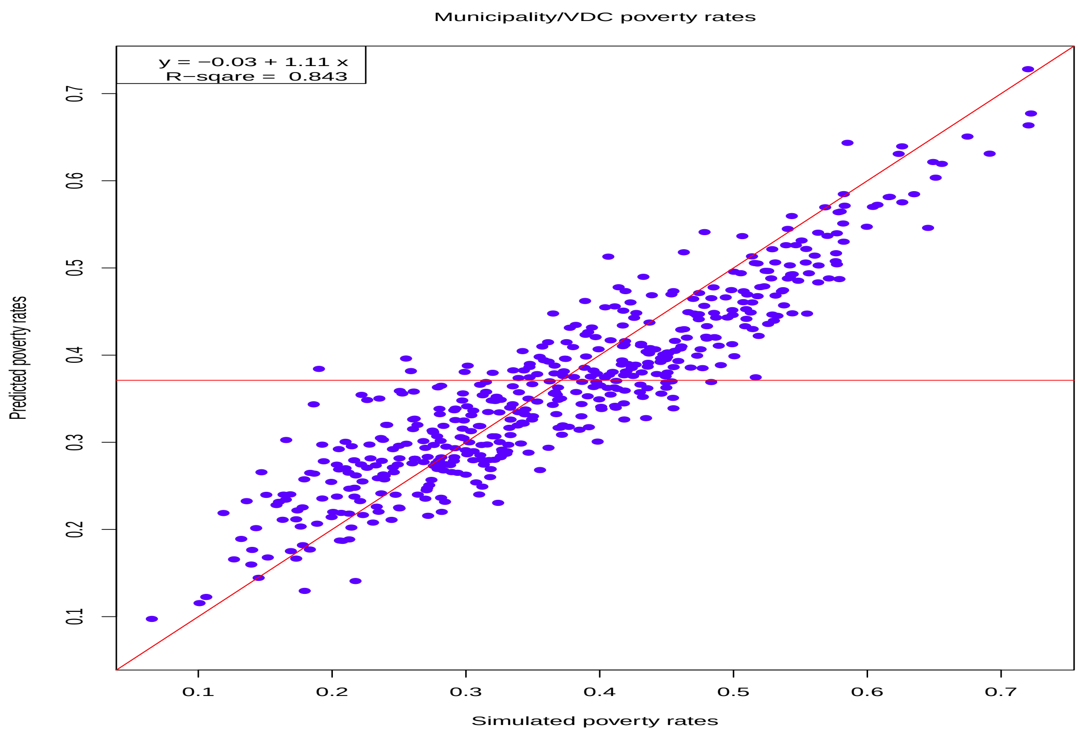


Figure 4.10: Poverty incidences in the simulation study by municipality/VDC (Noisy model, Mountains stratum)

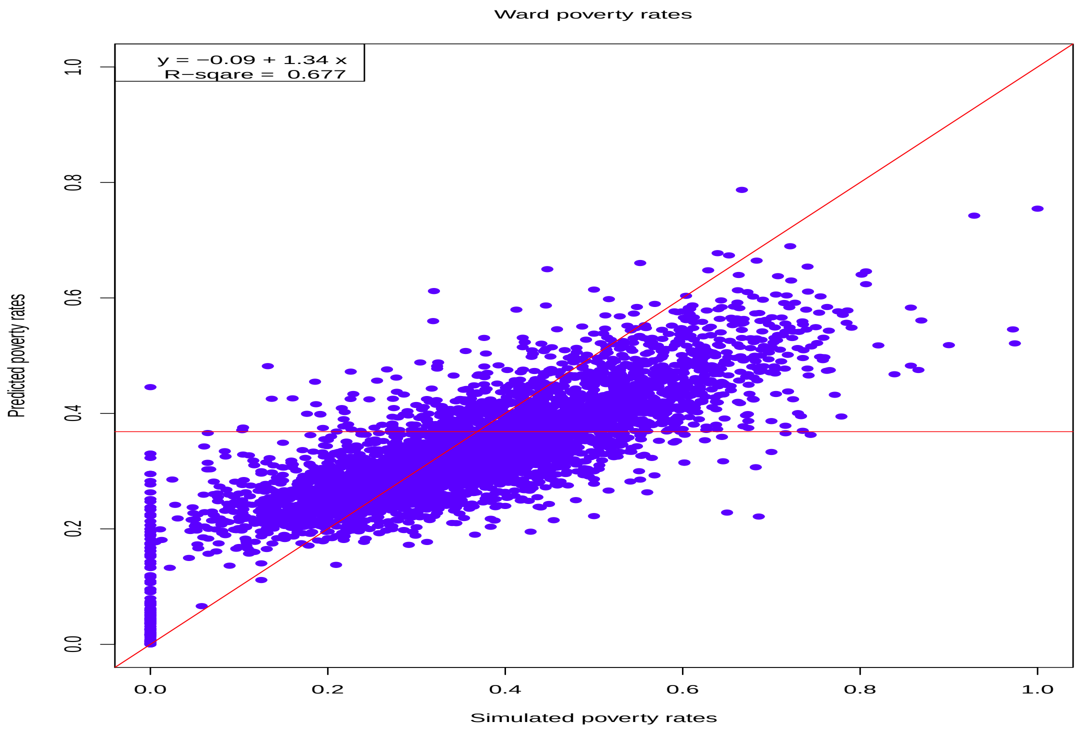


Figure 4.11: Poverty incidences in the simulation study by ward (Noiseless model, Mountains stratum)

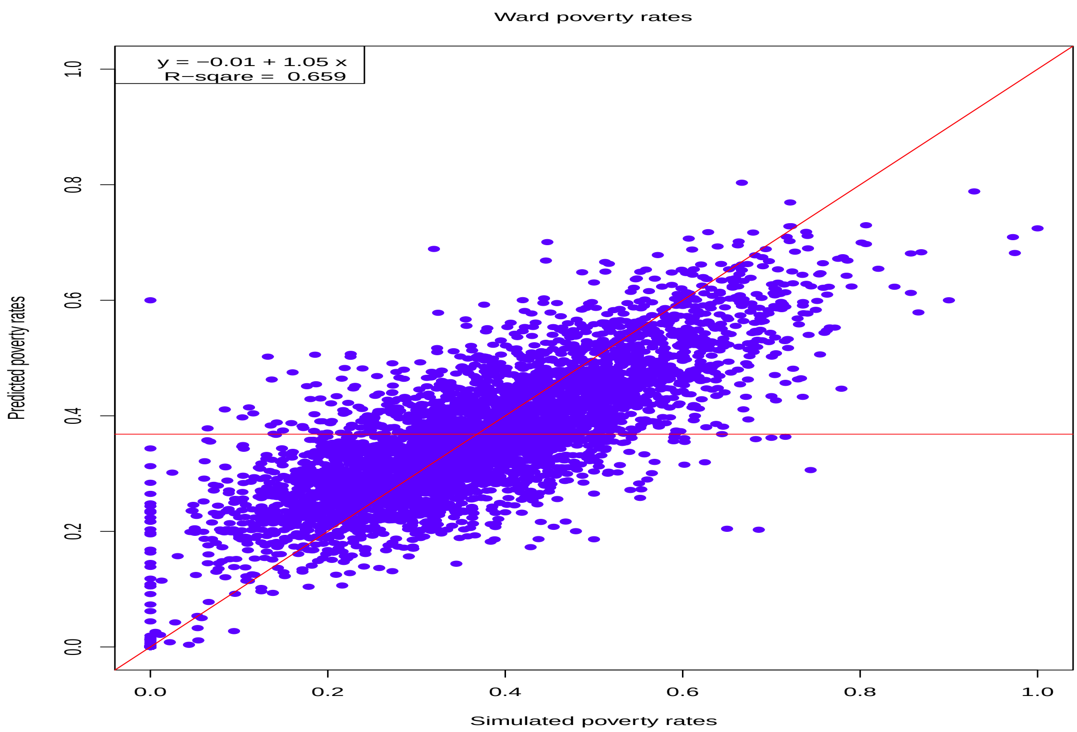


Figure 4.12: Poverty incidences in the simulation study by ward (Noisy model, Mountains stratum)

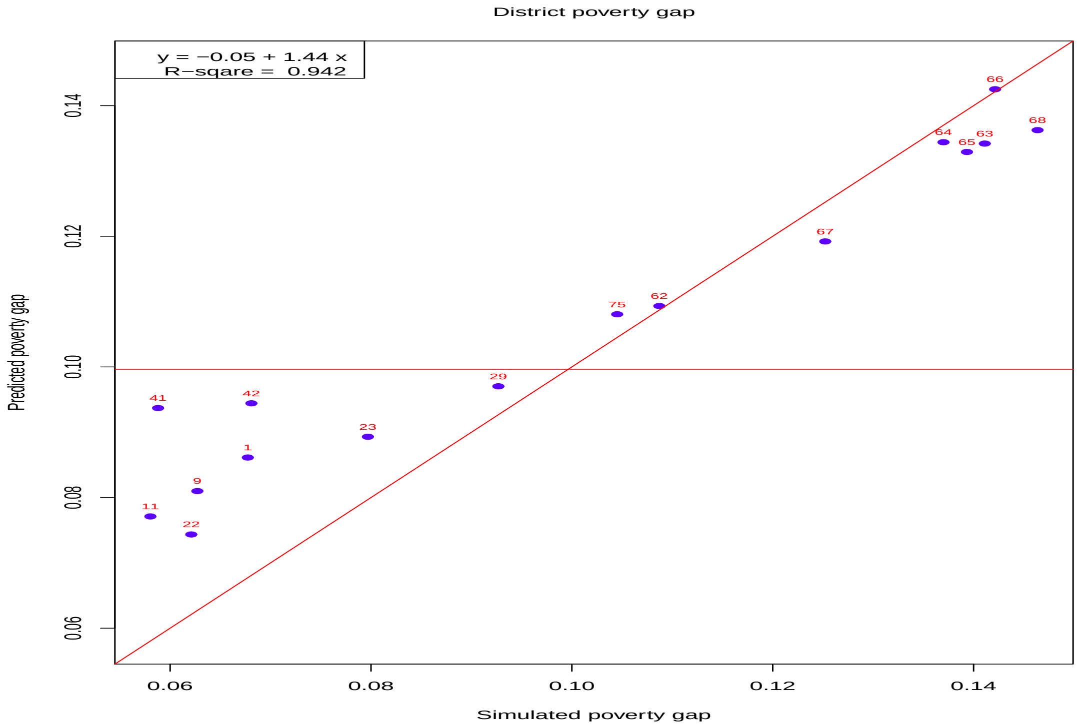


Figure 4.13: Poverty gaps in the simulation study by district (Noiseless model, Mountains stratum)

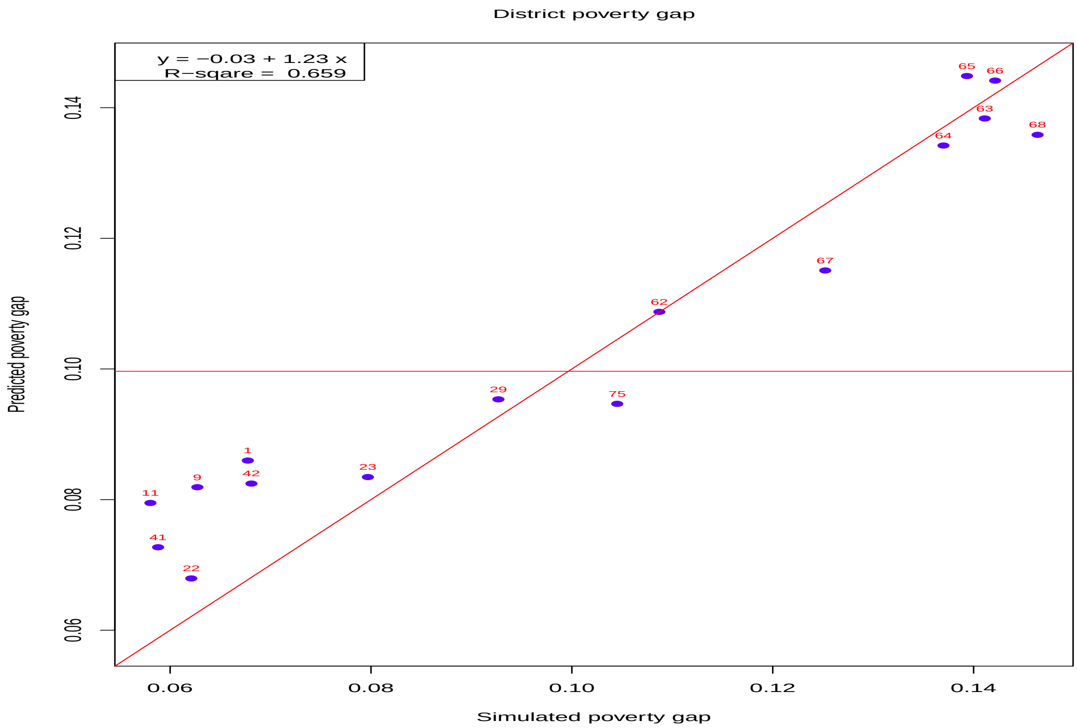


Figure 4.14: Poverty gaps in the simulation study by district (Noisy model, Mountains stratum)

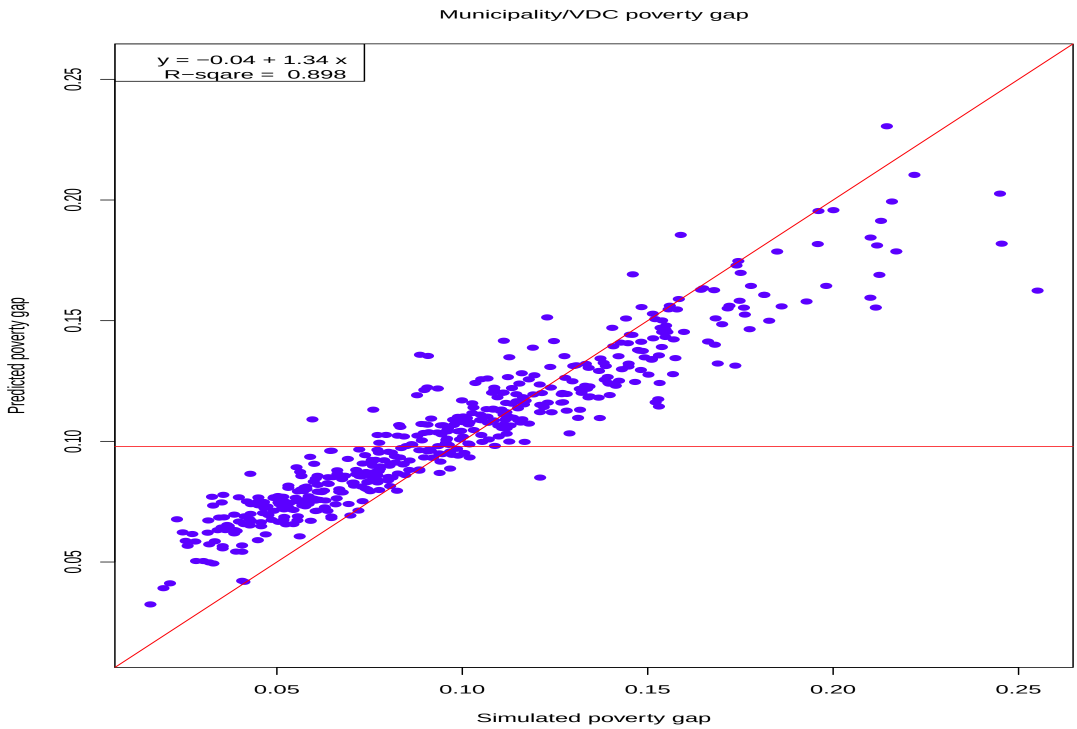


Figure 4.15: Poverty gaps in the simulation study by municipality/VDC (Noiseless model, Mountains stratum)

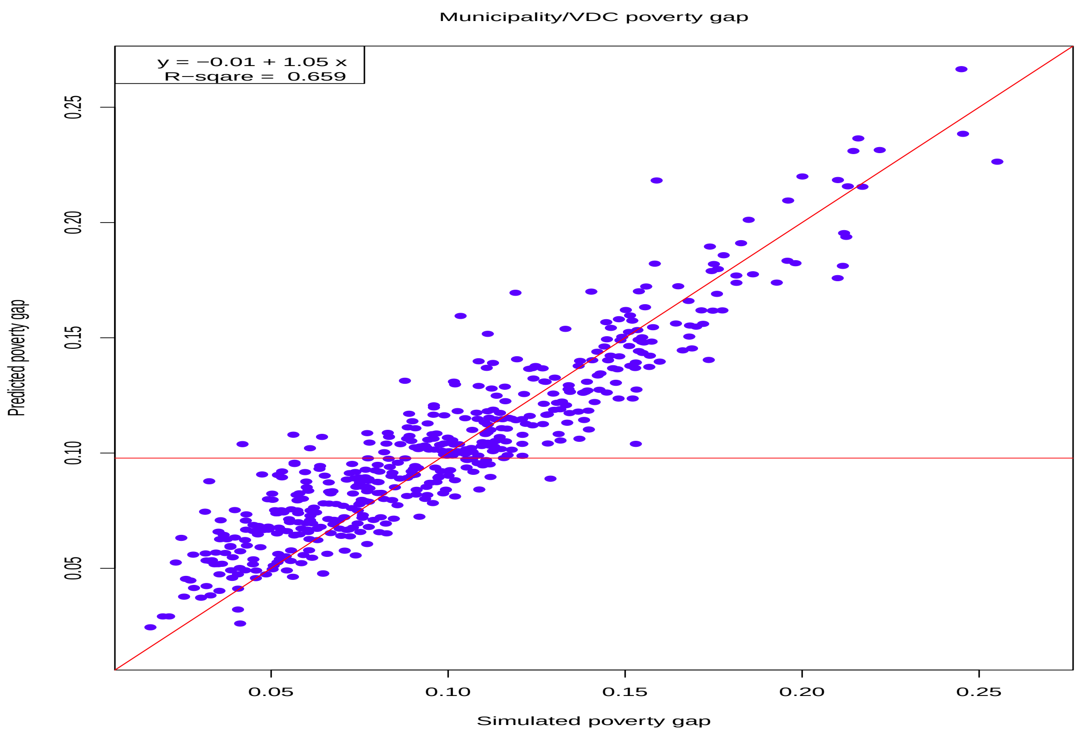


Figure 4.16: Poverty gaps in the simulation study by municipality/VDC (Noisy model, Mountains stratum)

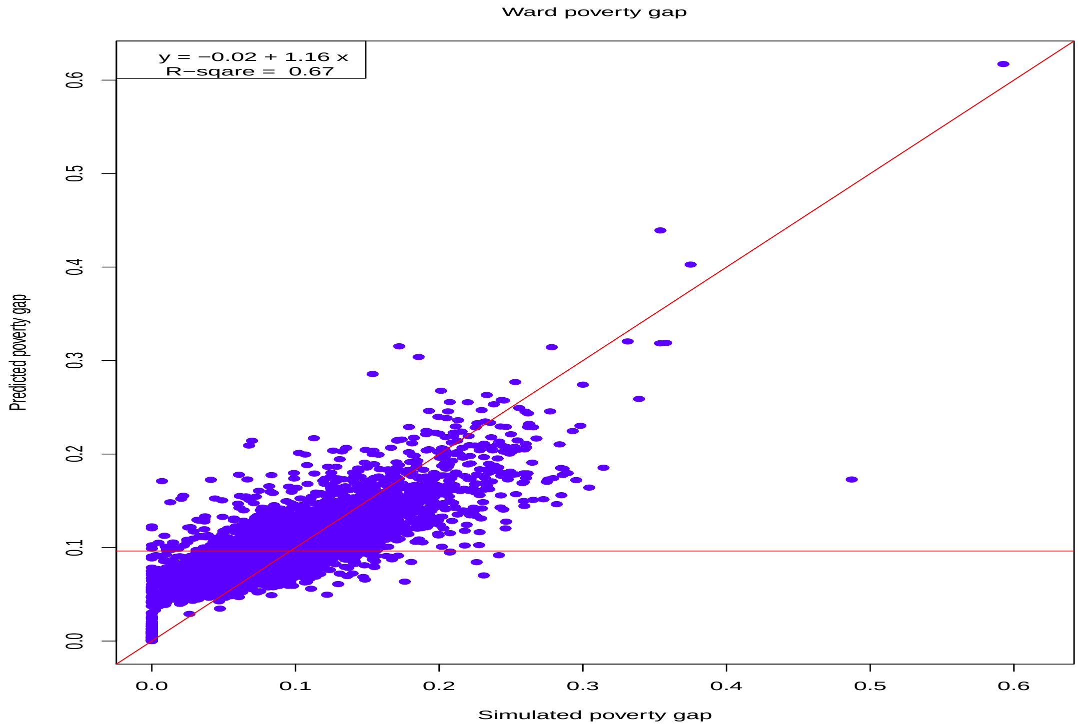


Figure 4.17: Poverty gaps in the simulation study by ward (Noiseless model, Mountains stratum)

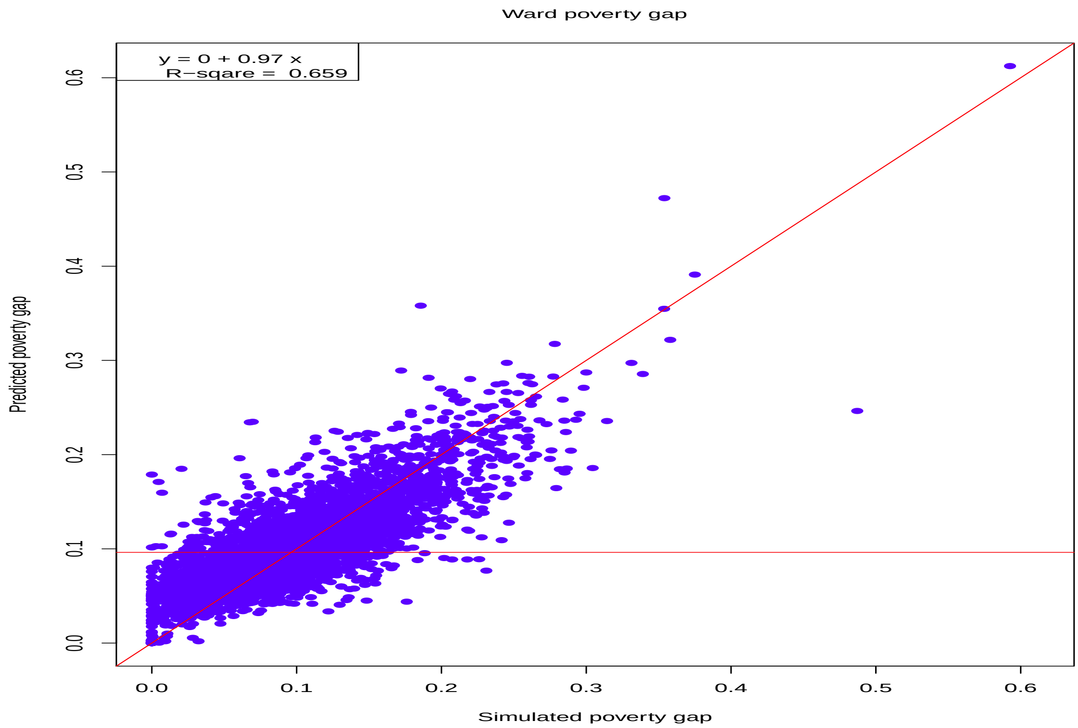


Figure 4.18: Poverty gaps in the simulation study by ward (Noisy model, Mountains stratum)

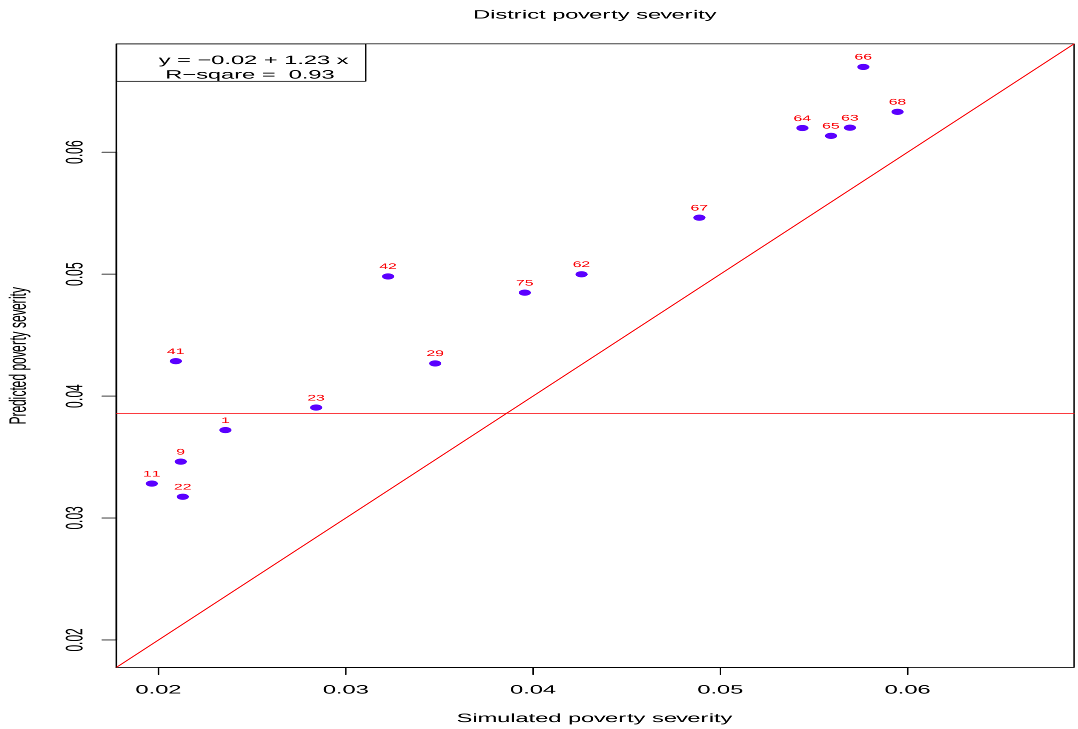


Figure 4.19: Poverty severities in the simulation study by district (Noiseless model, Mountains stratum)

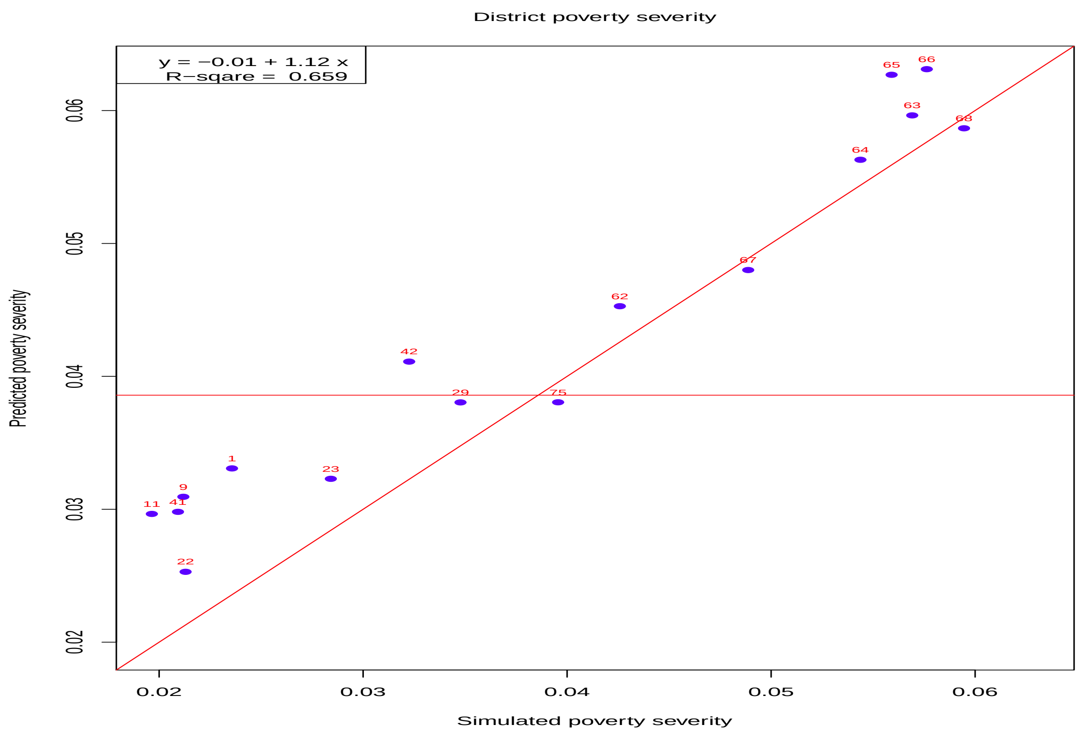


Figure 4.20: Poverty severities in the simulation study by district (Noisy model, Mountains stratum)

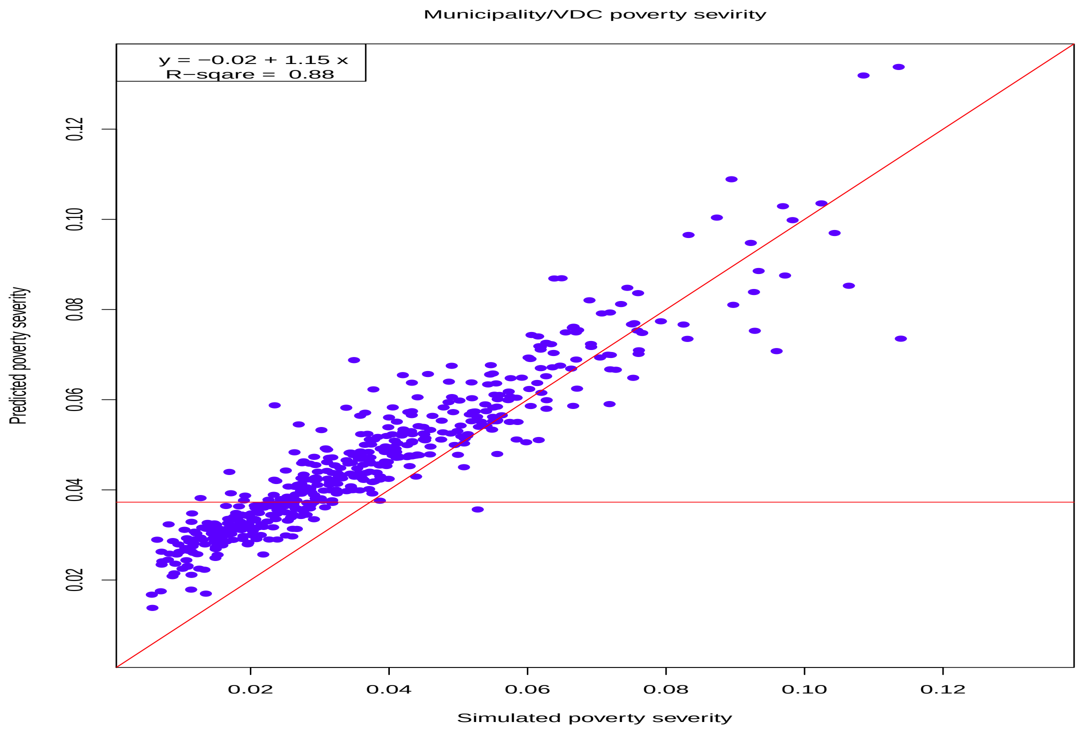


Figure 4.21: Poverty severities in the simulation study by municipality/VDC (Noiseless model, Mountains stratum)

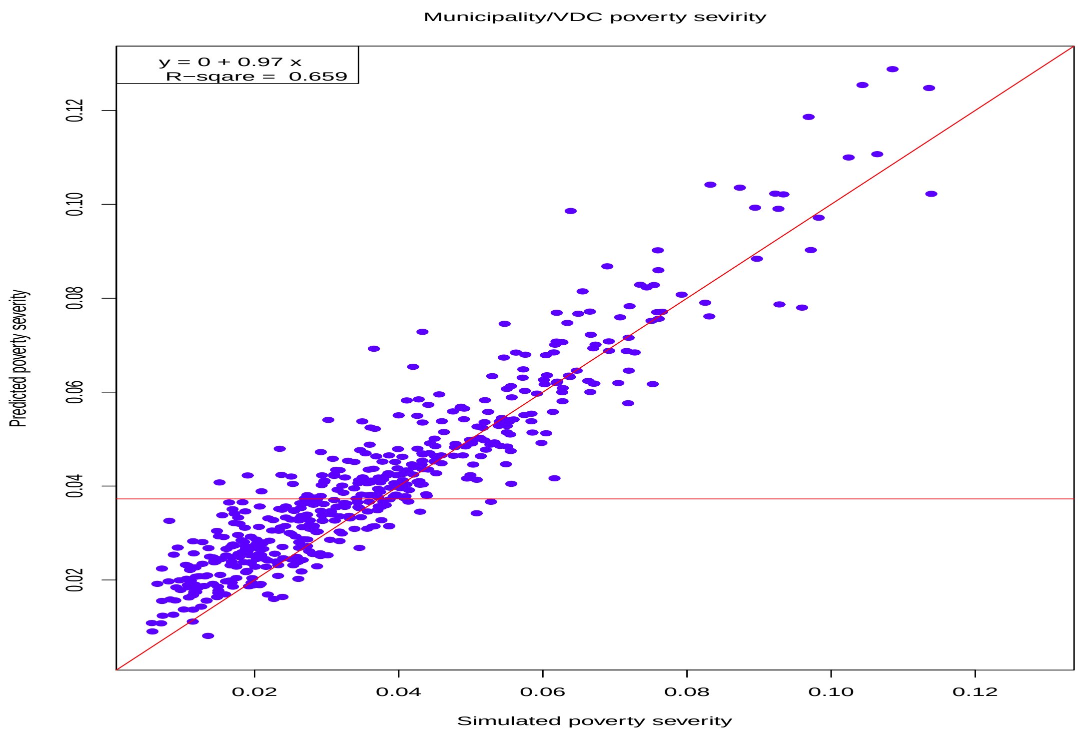


Figure 4.22: Poverty severities in the simulation study by municipality/VDC (Noisy model, Mountains stratum)

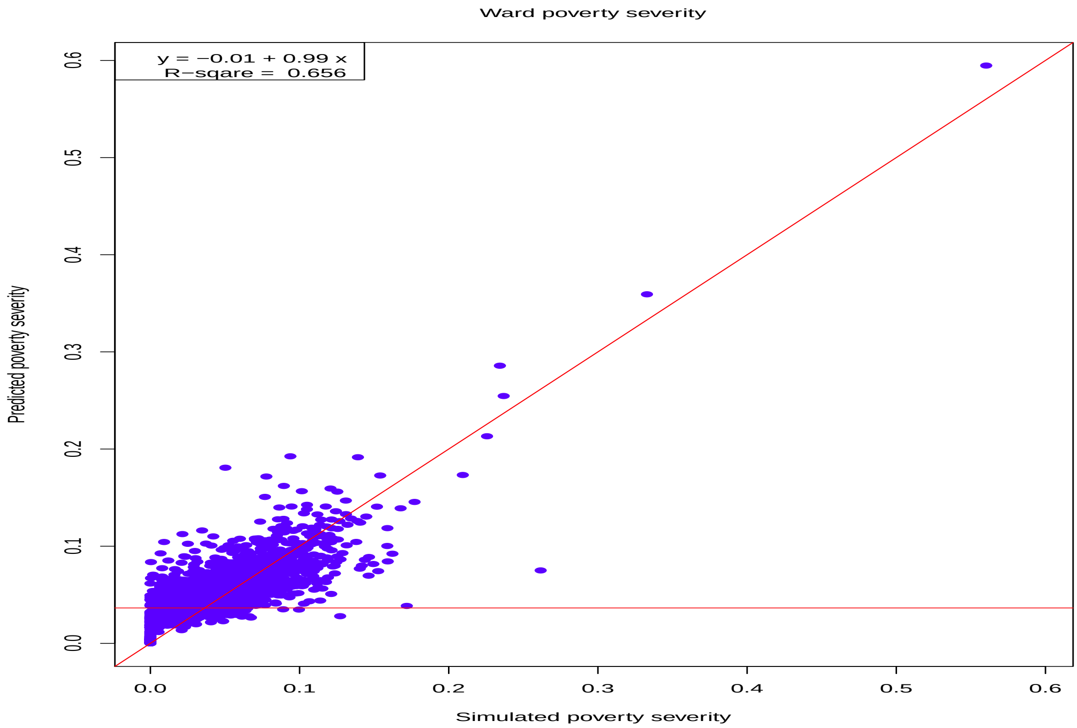


Figure 4.23: Poverty severities in the simulation study by ward (Noiseless model, Mountains stratum)

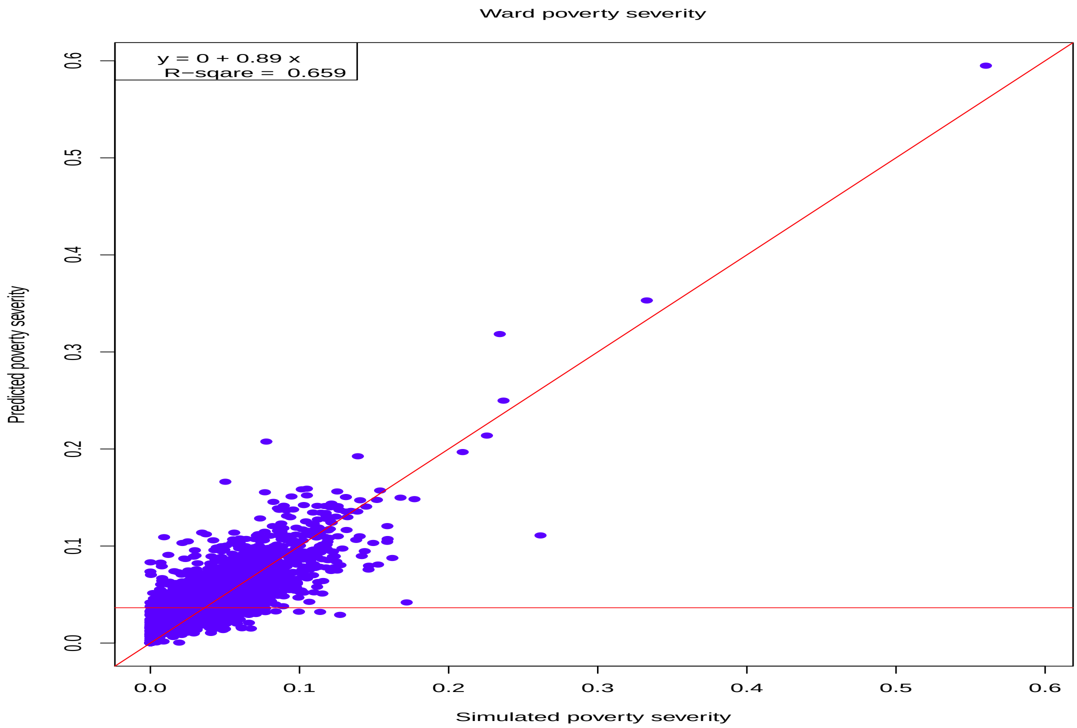


Figure 4.24: Poverty severities in the simulation study by ward (Noisy model, Mountains stratum)

Appendix A

Propriety of the posterior distributions

We conjecture that with some conditions the posterior densities under the noiseless model and noisy model are proper.

Theorem A.1

The joint posterior density for the generalized gamma model with random area effects $\pi(\boldsymbol{\nu}, \boldsymbol{\beta}, \alpha, \gamma, \sigma^2|\boldsymbol{y})$ *is proper if the design matrix is full rank and*

$$\int_{\boldsymbol{\beta}} e^{-\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\left(\alpha \boldsymbol{x'_{ij}}\boldsymbol{\beta} + y_{ij}^{\gamma} e^{-\boldsymbol{x'_{ij}}\boldsymbol{\beta}} - \frac{\sigma^2}{2}\left(y_{ij}^{\gamma} e^{-\boldsymbol{x'_{ij}}\boldsymbol{\beta}} - \alpha\, n_i\right)^2\right)} d\boldsymbol{\beta} \,<\, \infty.$$

Proof. We need to show

$$\int_{\sigma^2}\int_{\gamma}\int_{\alpha}\int_{\boldsymbol{\beta}}\int_{\boldsymbol{\nu}} \pi(\boldsymbol{\nu}, \boldsymbol{\beta}, \alpha, \gamma, \sigma^2|\boldsymbol{y})\, d\boldsymbol{\nu}\, d\boldsymbol{\beta}\, d\alpha\, d\gamma\, d\sigma^2 \,<\infty.$$

Let

$$\mathrm{T} = \int_{\sigma^2}\int_{\gamma}\int_{\alpha}\int_{\boldsymbol{\beta}}\int_{\boldsymbol{\nu}} \frac{e^{-R\gamma}\gamma^{S-1}}{(1+\sigma^2)^2\,(1+\alpha)^2}\left(\frac{1}{\sigma^2}\right)^{\frac{\ell}{2}} \times \left(\gamma\,\frac{g^{\alpha\gamma-1}}{\Gamma(\alpha)}\right)^n e^{-\alpha\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\boldsymbol{x'_{ij}}\boldsymbol{\beta}}$$
$$\times e^{-\sum_{i=1}^{\ell}\left(\alpha n_i \nu_i + e^{-\nu_i}\sum_{j=1}^{n_i} y_{ij}^{\gamma} e^{-\boldsymbol{x'_{ij}}\boldsymbol{\beta}} + \frac{\nu_i^2}{2\sigma^2}\right)} d\boldsymbol{\nu}\, d\boldsymbol{\beta}\, d\alpha\, d\gamma\, d\sigma^2.$$

Using the relation $e^{-\nu_i} \geq 1 - \nu_i$, we have

$$\mathrm{T} \leq \int_{\sigma^2}\int_{\gamma}\int_{\alpha}\int_{\boldsymbol{\beta}}\int_{\boldsymbol{\nu}} \frac{e^{-R\gamma}\gamma^{S-1}}{(1+\sigma^2)^2\,(1+\alpha)^2}\left(\frac{1}{\sigma^2}\right)^{\frac{\ell}{2}} \times \left(\gamma\,\frac{g^{\alpha\gamma-1}}{\Gamma(\alpha)}\right)^n e^{-\alpha\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\boldsymbol{x'_{ij}}\boldsymbol{\beta}}$$
$$\times e^{-\sum_{i=1}^{\ell}\left(\alpha n_i \nu_i + (1-\nu_i)\sum_{j=1}^{n_i} y_{ij}^{\gamma} e^{-\boldsymbol{x'_{ij}}\boldsymbol{\beta}} + \frac{\nu_i^2}{2\sigma^2}\right)} d\boldsymbol{\nu}\, d\boldsymbol{\beta}\, d\alpha\, d\gamma\, d\sigma^2.$$

Integrate out $\boldsymbol{\nu}$, we get

$$\mathrm{T} \leq \int_{\sigma^2}\int_{\gamma}\int_{\alpha}\int_{\boldsymbol{\beta}} \frac{e^{-R\gamma}\gamma^{S-1}}{(1+\sigma^2)^2\,(1+\alpha)^2}\left(\gamma\,\frac{g^{\alpha\gamma-1}}{\Gamma(\alpha)}\right)^n e^{-\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\left(\alpha \boldsymbol{x'_{ij}}\boldsymbol{\beta} + y_{ij}^{\gamma} e^{-\boldsymbol{x'_{ij}}\boldsymbol{\beta}} - \frac{\sigma^2}{2}\left(y_{ij}^{\gamma} e^{-\boldsymbol{x'_{ij}}\boldsymbol{\beta}} - \alpha\, n_i\right)^2\right)} d\boldsymbol{\beta}\, d\alpha\, d\gamma\, d\sigma^2.$$

It is difficult to integrate out $\boldsymbol{\beta}$. So we assume that

$$\int_{\boldsymbol{\beta}} e^{-\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\left(\alpha \boldsymbol{x'_{ij}}\boldsymbol{\beta} + y_{ij}^{\gamma} e^{-\boldsymbol{x'_{ij}}\boldsymbol{\beta}} - \frac{\sigma^2}{2}\left(y_{ij}^{\gamma} e^{-\boldsymbol{x'_{ij}}\boldsymbol{\beta}} - \alpha\, n_i\right)^2\right)} d\boldsymbol{\beta} \;=\; A(\alpha,\gamma,\sigma^2) \;<\; \infty.$$

where $A(\alpha,\gamma,\sigma^2)$, is the remaining normalizing constant after integrating out $\boldsymbol{\beta}$. Then

$$\mathrm{T} \leq \int_{\sigma^2}\int_{\gamma}\int_{\alpha} \frac{e^{-R\gamma}\gamma^{S-1}}{(1+\sigma^2)^2\,(1+\alpha)^2} \times \left(\gamma\,\frac{g^{\alpha\gamma-1}}{\Gamma(\alpha)}\right)^n \times A(\alpha,\gamma,\sigma^2)\; d\alpha\, d\gamma\, d\sigma^2.$$

We replace α by α^* in the expression $\gamma\,\frac{g^{\alpha\gamma-1}}{\Gamma(\alpha)}$ by its maximum since it is log-concave with respect to α

$$\gamma\,\frac{g^{\alpha\gamma-1}}{\Gamma(\alpha)} \leq \gamma\,\frac{g^{\alpha^*\gamma-1}}{\Gamma(\alpha^*)}$$

We assume that $\left(\gamma\,\frac{g^{\alpha^*\gamma-1}}{\Gamma(\alpha^*)}\right)^n \times A(\alpha,\gamma,\sigma^2) < \infty$. Since we have proper priors for α,γ and σ^2, we have the proper posterior. $\square$

Theorem A.2

The joint posterior density for the mixture of two generalized gamma (GB2) model with random area effects $\pi(\alpha,\boldsymbol{\beta},\gamma,\boldsymbol{\nu},\sigma^2|\boldsymbol{y})$ is proper if the design matrix is full rank and

$$\int_{\boldsymbol{\beta}} e^{-\sum_{i=1}^{\ell}\left[\alpha\sum_{j=1}^{n_i}\boldsymbol{x'_{ij}}\boldsymbol{\beta} + \sum_{j=1}^{n_i}\left(y_{ij}\, e^{-\boldsymbol{x'_{ij}}\boldsymbol{\beta}}\right)^{\frac{2(\alpha+1)}{a}} - \sigma^2\left\{\frac{2(\alpha+1)}{a}\sum_{j=1}^{n_i}\left(y_{ij}\, e^{-\boldsymbol{x'_{ij}}\boldsymbol{\beta}}\right)^{\frac{2(\alpha+1)}{a}} - \alpha\, n_i\right\}^2\right]} d\boldsymbol{\beta} \;<\; \infty.$$

Proof. We need to show

$$\int_{\sigma^2}\int_{\gamma}\int_{\alpha}\int_{\boldsymbol{\beta}}\int_{\boldsymbol{\nu}} \pi(\boldsymbol{\nu},\boldsymbol{\beta},\alpha,\gamma,\sigma^2|\boldsymbol{y})\; d\boldsymbol{\nu}\, d\boldsymbol{\beta}\, d\alpha\, d\gamma\, d\sigma^2 \;<\infty.$$

Let

$$\mathrm{T} = \int_{\sigma^2}\int_{\gamma}\int_{\alpha}\int_{\boldsymbol{\beta}}\int_{\boldsymbol{\nu}} \frac{e^{-R\gamma}\gamma^{S-1}}{(1+\sigma^2)^2\,(1+\alpha)^2}\left[\frac{\gamma\, g^{\alpha-1}}{\mathrm{B}(\frac{\alpha}{\gamma},\frac{\alpha+2}{\gamma})}\right]^n \left(\frac{1}{\sigma^2}\right)^{\frac{\ell}{2}} \times e^{-\alpha\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\boldsymbol{x'_{ij}}\boldsymbol{\beta}}$$

$$\times \prod_{i=1}^{\ell}\left[\frac{e^{-\left(\alpha\, n_i\nu_i + \frac{\nu_i^2}{2\sigma^2}\right)}}{\prod_{j=1}^{n_i}\left(1+\left[y_{ij}\, e^{-(\boldsymbol{x'_{ij}}\boldsymbol{\beta}+\nu_i)}\right]^{\gamma}\right)^{\frac{2(\alpha+1)}{\gamma}}}\right] d\boldsymbol{\nu}\, d\boldsymbol{\beta}\, d\alpha\, d\gamma\, d\sigma^2.$$

For $a > 0$, large enough, $e^{-x} \geq \frac{1}{(1+x)^a}$ for almost all $x > 0$. We can choose a such that

$|e^{-x} - \frac{1}{(1+x)^a}| < \xi$ for any $\xi > 0$. So

$$\mathrm{T} \leq \int_{\sigma^2} \int_\gamma \int_\alpha \int_{\boldsymbol{\beta}} \int_{\boldsymbol{\nu}} \frac{e^{-R\gamma} \gamma^{S-1}}{(1+\sigma^2)^2\,(1+\alpha)^2} \left[\frac{\gamma\, g^{\alpha-1}}{\mathrm{B}(\frac{\alpha}{\gamma}, \frac{\alpha+2}{\gamma})} \right]^n \left(\frac{1}{\sigma^2}\right)^{\frac{\ell}{2}} \times e^{-\alpha \sum_{i=1}^{\ell} \sum_{j=1}^{n_i} \boldsymbol{x'_{ij}\beta}}$$
$$\times \prod_{i=1}^{\ell} \left[e^{-\left(\alpha\, n_i \nu_i + \frac{\nu_i^2}{2\sigma^2}\right)} \prod_{j=1}^{n_i} \left\{ e^{-\left(y_{ij}\, e^{-(\boldsymbol{x'_{ij}\beta}+\nu_i)}\right)^{\frac{2(\alpha+1)}{a}}} \right\} \right] .d\boldsymbol{\nu}\, d\boldsymbol{\beta}\, d\alpha\, d\gamma\, d\sigma^2.$$

Now, using $e^{-\frac{2(\alpha+1)}{a}\nu_i} \geq 1 - \frac{2(\alpha+1)}{a}\nu_i$

$$\mathrm{T} \leq \int_{\sigma^2} \int_\gamma \int_\alpha \int_{\boldsymbol{\beta}} \int_{\boldsymbol{\nu}} \frac{e^{-R\gamma} \gamma^{S-1}}{(1+\sigma^2)^2\,(1+\alpha)^2} \left[\frac{\gamma\, g^{\alpha-1}}{\mathrm{B}(\frac{\alpha}{\gamma}, \frac{\alpha+2}{\gamma})} \right]^n \left(\frac{1}{\sigma^2}\right)^{\frac{\ell}{2}} \times e^{-\alpha \sum_{i=1}^{\ell} \sum_{j=1}^{n_i} \boldsymbol{x'_{ij}\beta}}$$
$$\times e^{-\sum_{i=1}^{\ell}\left(\alpha\, n_i \nu_i + \frac{\nu_i^2}{2\sigma^2}\right)}\; e^{-\sum_{i=1}^{\ell} \sum_{j=1}^{n_i} \left(1 - \frac{2(\alpha+1)}{a}\nu_i\right)\left(y_{ij}\, e^{-\boldsymbol{x'_{ij}\beta}}\right)^{\frac{2(\alpha+1)}{a}}}\, d\boldsymbol{\nu}\, d\boldsymbol{\beta}\, d\alpha\, d\gamma\, d\sigma^2.$$

We have $\nu_i, i = 1, \cdots, \ell$ independent. Integrating out $\boldsymbol{\nu}$, we have

$$\mathrm{T} \leq \int_{\sigma^2} \int_\gamma \int_\alpha \int_{\boldsymbol{\beta}} \frac{e^{-R\gamma} \gamma^{S-1}}{(1+\sigma^2)^2\,(1+\alpha)^2} \left[\frac{\gamma\, g^{\alpha-1}}{\mathrm{B}(\frac{\alpha}{\gamma}, \frac{\alpha+2}{\gamma})} \right]^n \times e^{-\alpha \sum_{i=1}^{\ell} \sum_{j=1}^{n_i} \boldsymbol{x'_{ij}\beta}}$$
$$\times\, e^{-\sum_{i=1}^{\ell}\left[\alpha \sum_{j=1}^{n_i} \boldsymbol{x'_{ij}\beta} + \sum_{j=1}^{n_i}\left(y_{ij}\, e^{-\boldsymbol{x'_{ij}\beta}}\right)^{\frac{2(\alpha+1)}{a}} - \sigma^2 \left\{ \frac{2(\alpha+1)}{a} \sum_{j=1}^{n_i}\left(y_{ij}\, e^{-\boldsymbol{x'_{ij}\beta}}\right)^{\frac{2(\alpha+1)}{a}} - \alpha\, n_i \right\}^2 \right]}\, d\boldsymbol{\beta}\, d\alpha\, d\gamma\, d\sigma^2.$$

It is difficult to integrate out $\boldsymbol{\beta}$. So we assume that

$$\int_{\boldsymbol{\beta}} e^{-\sum_{i=1}^{\ell}\left[\alpha \sum_{j=1}^{n_i} \boldsymbol{x'_{ij}\beta} + \sum_{j=1}^{n_i}\left(y_{ij}\, e^{-\boldsymbol{x'_{ij}\beta}}\right)^{\frac{2(\alpha+1)}{a}} - \sigma^2 \left\{ \frac{2(\alpha+1)}{a} \sum_{j=1}^{n_i}\left(y_{ij}\, e^{-\boldsymbol{x'_{ij}\beta}}\right)^{\frac{2(\alpha+1)}{a}} - \alpha\, n_i \right\}^2 \right]}\, d\boldsymbol{\beta} \;=\; A(\alpha, \sigma^2) \;<\; \infty.$$

where $A(\alpha, \sigma^2)$, is the remaining normalizing constant after integrating out $\boldsymbol{\beta}$. Then

$$\mathrm{T} \leq \int_{\sigma^2} \int_\gamma \int_\alpha \frac{e^{-R\gamma} \gamma^{S-1}}{(1+\sigma^2)^2\,(1+\alpha)^2} \left[\frac{\gamma\, g^{\alpha-1}}{\mathrm{B}(\frac{\alpha}{\gamma}, \frac{\alpha+2}{\gamma})} \right]^n \times A(\alpha, \sigma^2)\; d\alpha\, d\gamma\, d\sigma^2.$$

We replace α by α^* in the expression $\frac{\gamma\, g^{\alpha-1}}{\mathrm{B}(\frac{\alpha}{\gamma}, \frac{\alpha+2}{\gamma})}$ by its maximum since it is log-concave with respect to α

$$\frac{\gamma\, g^{\alpha-1}}{\mathrm{B}(\frac{\alpha}{\gamma}, \frac{\alpha+2}{\gamma})} \leq \frac{\gamma\, g^{\alpha^*-1}}{\mathrm{B}(\frac{\alpha^*}{\gamma}, \frac{\alpha^*+2}{\gamma})}$$

We assume that $\frac{\gamma\, g^{\alpha^*-1}}{\mathrm{B}(\frac{\alpha^*}{\gamma}, \frac{\alpha^*+2}{\gamma})} \times A(\alpha, \sigma^2) < \infty$. Since we have proper priors for α, γ and σ^2, we have the proper posterior. $\square$

Appendix B

Bayesian Bootstrap

The Bayesian bootstrap (BB) is the Bayesian analogue of the classical bootstrap. Suppose we have observed n sample values $\nu_1, \cdots, \nu_n$, which are viewed as n $i.i.d$ samples of the random variable X. Let us assume there are c distinct values of the samples $\tilde{\nu_1}, \tilde{\nu_2}, \cdots, \tilde{\nu_c}$, in the n samples with corresponding frequencies $f_1, \cdots, f_c$. For these distinct categories with some unknown probabilities $p_1, p_2, \cdots, p_c$, we assume that $f_1, \cdots, f_c$ follow a multinomial distribution. With the Dirichlet($\mathbf{0}$) as a prior distribution for parameter $\boldsymbol{p}$, we can write the Bayesian Bootstrap as follows

$$f_1, \cdots, f_c | \boldsymbol{p} \sim \text{Multinomial}(n, \boldsymbol{p}), \quad \pi(\boldsymbol{p}) \sim \text{Dirichlet}(\mathbf{0}).$$

The posterior distribution function of $\boldsymbol{p}|\boldsymbol{f}$ given by

$$\pi(\boldsymbol{p}|\boldsymbol{f}) \propto \prod_{i=1}^{c} p_i^{f_i - 1},$$

is also a Dirichlet distribution, $\pi(\boldsymbol{p}|\boldsymbol{f}) \sim \text{Dirichlet}(\boldsymbol{f})$. In the Bayesian bootstrap, we draw the probabilities from the above Dirichlet posterior distribution, and then draw an indicator of the classes, $\tilde{\nu_1}, \tilde{\nu_2}, \cdots, \tilde{\nu_c}$, from the multinomial distribution, $I|n, \mathbf{p} \sim \text{Multinomial}\,(\mathbf{1}, \mathbf{p})$, to get one of the distinct observation $\tilde{\nu_1}, \tilde{\nu_2}, \cdots, \tilde{\nu_c}$. This procedure can be repeated as many times as needed.

Appendix C

Response and Covariates

S.N	Variable name	Description
1	y	Real per capita consumption per year (Response)
2	hhsize	Household size
3	skids6	Proportion of kids aged 0 - 6 in the household
4	skids714	Proportion of kids aged 7 - 14 in the household
5	remtab	Abroad migrant
6	hutype3	House temporary
7	huwon2	House owned
8	ckfuel3w	Proportion of households with cooking fule LP/gas in Ward
9	pflandv	Proportion of household with land-owning females in VDC
10	pschv	Proportion of kids 6-16 attending school in VDC

Appendix D

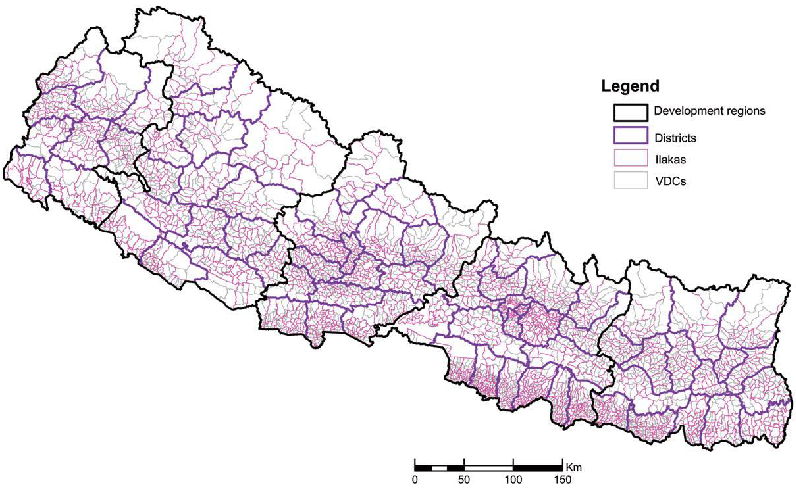
Stratum Names

Table D.1: Stratum names in NLSS-II

Stratum	Stratum Name
1	Mountains
2	Kathmandu Urban Valley
3	Urban Hills
4	Rural Hills
5	Urban Terai
6	Rural Tarai

Appendix E

Political map of Nepal and PSUs in NLSS-II



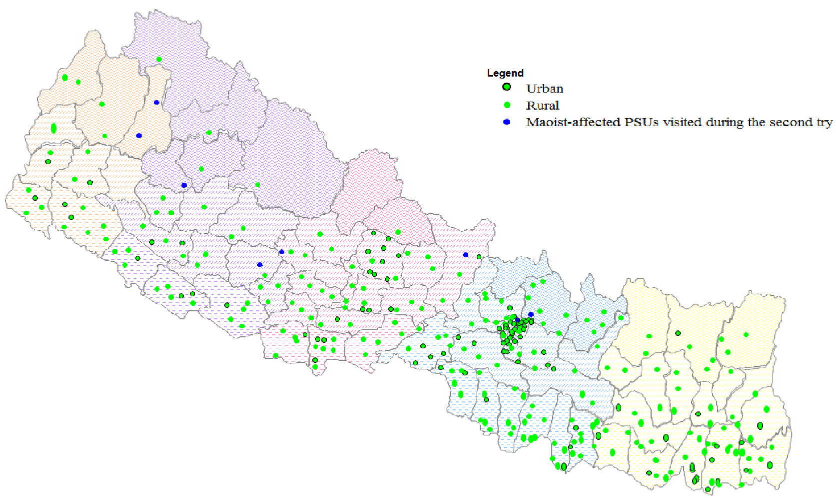


Figure E.1: Political municipality/Village development committee map of Nepal and NLSS-II PSUs

Bibliography

[1] Battese, G. E., Harter, R. M., and Fuller, W. A. (1988), *An error-components model for prediction of county crop areas using survey and satellite data*, Journal of the American Statistical Association 83.401, 28-36.

[2] Berliner, L. M (1996), *Hierarchical Bayesian time series models*, Maximum entropy and Bayesian methods, 15-22, Springer.

[3] Besag, J. (1974), *Spatial interaction and the statistical analysis of lattice systems*, Journal of the Royal Statistical Society. Series B (Methodological), 192-236.

[4] Box, G. E. (1980), *Sampling and Bayes' Inference in Scientific Modelling and Robustness*, Journal of the Royal Statistical Society, Series A (General), 4, 383-430, Royal Statistical Society, Wiley, ISSN 00359238.

[5] Central Bureau of Statistics, Nepal, *Nepal living standards survey report 2003/04*, Vol 1.

[6] Central Bureau of Statistics, Nepal, *Nepal living standards survey report 2003/04*, Vol 2.

[7] Central Bureau of Statistics, Nepal, *Poverty trends in Nepal (1995-96 and 2003-04)*, CBS Nepal.

[8] Cressie, N., and Wikle, C. K. (2015), *Statistics for spatio-temporal data*, John Wiley & Sons.

[9] Cummins, J. D., Dionne, G., McDonald, J. B., and Pritchett, B. M. (1990), *Applications of the GB2 family of distributions in modeling insurance loss processes*, Insurance: Mathematics and Economics, 9(4), 257-272.

[10] Dutta, K., and Perry, J. (2006), *A tale of tails: an empirical analysis of loss distribution models for estimating operational risk capital.*.

[11] Dong, A. X., and Chan, J. S. K. (2013), *Bayesian analysis of loss reserving using dynamic models with generalized beta distribution*, Insurance: Mathematics and Economics, 53, 2, 355-365.

[12] Elbers, C., Lanjouw, J. O., and Lanjouw, P. (2003), *Micro-level estimation of poverty and inequality*, Econometrica, 71, 1, 355-364, Wiley Online Library.

[13] Fabrizi, E., Rosaria, M. and Trivisano, C. (2016), *Bayesian beta regression models for the estimation of poverty and inequality parameters in small area*, In *Analysis of Poverty Data by Small Area Estimation*, Ed. Pratesi, M., Wiley, 299-314.

[14] Feng, C., Wang, H., Lu, N., and Tu, X. M. (2013), *Log transformation: application and interpretation in biomedical research*, Statistics in medicine, 32, 230-239.

[15] Feng, C., Wang, H., Lu, N., Chen, T., He, H., Lu, Y., and Tu, X. M. (2014), *Log-transformation and its implications for data analysis*, Shanghai archives of psychiatry, 26, 105-109.

[16] Foster, J., Greer, J., and Thorbecke, E. (1984), *A Class of Decomposable Poverty Measures*, Econometrica, 3, 761-766, Wiley, Econometric Society, 52.

[17] Fujii, T. (2004), *Commune-level estimation of poverty measures and its application in Cambodia*, UNU-WIDER Research Paper 2004/48.

[18] Gelfand, A. E., Dey, D. K., and Chang, H. (1992), *Model determination using predictive distributions with implementation via sampling-based methods*, DTIC Document.

[19] Ghosh, M. and Rao, J. N. K. (1994), *Small Area Estimation: An Appraisal*, Statistical Science, 9, 55-76.

[20] Graf, M. and Nedyalkova, D. (2014), *Modeling of income and indicators of poverty and social exclusion using the generalized beta distribution of the second kind*, Review of Income and Wealth, 60, 821-842.

[21] Hager, H. W., and Bain, L. J. (1970), *Inferential procedures for the generalized gamma distribution*, Journal of the American Statistical Association, 65(332), 1601-1609.

[22] Haslett, S. J., and Jones, G. (2006), *Small area estimation of poverty, caloric intake and malnutrition in Nepal*, Nepal Central Bureau of Statistics/World Food Programme, United Nations/World Bank, Kathmandu, 184, 999337018-5.

[23] Haslett, S. J., Jones, G., and Sefton, A. (2012), *Small-Area Estimation of Poverty and Malnutrition in Cambodia*, The National Institute of Statistics, Cambodia in collaboration with The United Nations World Food Programme.

[24] Haughton, D., and Haughton, J. (2011), *Living standards analytics: Development through the lens of household survey data*, Springer Science & Business Media.

[25] He, Z., and Sun, D. (2000), *Hierarchical Bayes estimation of hunting success rates with spatial correlations*, Biometrics, 56(2), 360-367.

[26] Huang, P. H., and Hwang, T. Y., (2006), *On new moment estimation of parameters of the generalized gamma distribution using it's characterization*, Taiwanese journal of Mathematics, 1083-1093.

[27] Hwang, T. Y., and Hu, C. Y., (1999), *On a characterization of the gamma distribution: The independence of the sample mean and the sample coefficient of variation*,Annals of the Institute of Statistical Mathematics, 51(4), 749-753.

[28] Hwang, T. Y., and Hu, C. Y., (2000), *On some characterizations of population distributions*, Taiwanese Journal of Mathematics, 427-437.

[29] Kerman, S. C., and McDonald, J. B. (2015), *Skewness-kurtosis bounds for EGB1, EGB2, and Special Cases*, Communications in Statistics theory and Methods, 44(18), 3857-3864.

[30] Khodabin, M., and Ahmadabadi, A. (2010), *Some properties of generalized gamma distribution*, Mathematical Sciences, 4(1), 9-28.

[31] Lienhard, J. H., and Meyer, P. L. (1967), *A physical basis for the generalized gamma distribution*, Quarterly of applied mathematics, 25(3), 330-334.

[32] **Manandhar, B.** (2009), *Statistics of Newar and Nepalmandal Territory: Analyzing National Population Censuses Data 1991-2001 of Newaa Civilization*, Ilohan Publication, Kathmandu Nepalmandal, ISBN: 978-9937-9015-2-9

[33] **Manandhar, B.** and Nandram, B. (2016), *Bayesian predictive inference for consumption data from small areas*, In JSM 2016 proceedings, Section on Bayesian Statistical Science, Alexandria, VA: American Statistical Association, 2437-2446.

[34] **Manandhar, B.** and Nandram, B. (2017a), *Hierarchical Bayesian models for size responses from small areas: An application to poverty estimation*, **(under preparation)**.

[35] **Manandhar, B.** and Nandram, B. (2017b), *Hierarchical Bayesian models for noisy size responses from small areas: An application to poverty estimation*, **(under preparation)**.

[36] McDonald, J. B. (1984), *Some generalized functions for the size distribution of income*, Econometrica: journal of the Econometric Society, 647-663.

[37] McDonald, J. B. and Xu, Y. J. (1995), *A generalization of the beta distribution with applications*, Journal of Econometrics, 66, 133-152.

[38] Molina, I., and Rao, J. N. K. (2010), *Small area estimation of poverty indicators*, Canadian Journal of Statistics, 38, 3, 369-385.

[39] Molina, I., Nandram, B., and Rao, J. N. K. (2014), *Small area estimation of general parameters with application to poverty indicators: A hierarchical Bayes approach*, The Annals of Applied Statistics, 8, 852-885.

[40] Mosimann, J. E. (1970), *Size allometry: size and shape variables with characterizations of the lognormal and generalized gamma distributions*, Journal of the American Statistical Association, 65(330), 930-945.

[41] Moura, F. A., and Migon, H. S. (2002), *Bayesian spatial models for small area estimation of proportions*, Statistical Modelling, 2(3), 183-201.

[42] Munoz, J. (2003), *Nepal living standards survey progress report, 2003/04* , Unpublished sample design report.

Author, A. A. (1996). Title of paper or manuscript. Unpublished manuscript.

[43] Nandram, B. and Sedransk, J and Pickle, L. W. (2000), *Bayesian analysis and mapping of mortality rates for chronic obstructive pulmonary disease*, Journal of the American Statistical Association, 95, 452, 1110-1118, Taylor & Francis Group.

[44] Nandram, B. and Choi, J. W. (2005), *Hierarchical Bayesian nonignorable nonresponse regression models for small areas: An application to the NHANES data*, Survey Methodology, 32, 1, 73-84, Statistics Canada.

[45] Nandram, B. and Choi, J. W. (2010), *A Bayesian analysis of body mass index data from small domains under nonignorable nonresponse and selection*, Journal of the American Statistical Association, 105, 120-135.

[46] Nandram, B., Fu, S., and **Manandhar, B.** (2017), *Bayesian logistic regression for small areas with numerous households*, **(under preparation)**.

[47] Nandram, B., Xu, Z., and **Manandhar, B.** (2017), *Bayesian inference of a finite population under length bias*, **(under preparation)**.

[48] Nandram, B. and Yin, J. (2016a), *A nonparametric Bayesian prediction interval for a finite population mean*, Journal of Statistical Computation and Simulation, 86(16), 3141-3157.

[49] Nandram, B. and Yin, J. (2016b), *Bayesian predictive inference under a Dirichlet process with sensitivity to the normal baseline*, Statistical Methodology, 28, 1-17.

[50] Nguyen, P. (2014), *Viet Nam household living standards surveys 2002-2008 in time and space: managerial aspects and multilevel models*, (Doctoral dissertation).

[51] Nguyen, P., Haughton, D., Hudson, I., and Boland, J. (2010), *Multilevel models ans small area estimation in the context of Vietnam living standards surveys*, 42emes Journes de Statistique.

[52] Peters, G. W., and Sisson, S. A. (2006), *Bayesian inference, Monte Carlo sampling and operational risk*, Journal of Operational Risk, 1(3), 27-50.

[53] Pettit, L. I. (1990), *The conditional predictive ordinate for the normal distribution*, Journal of the Royal Statistical Society. Series B (Methodological), 175-184, JSTOR.

[54] Prentice, R. L. (1974), *TA log gamma model and its maximum likelihood estimation*, Biometrika, 539-544.

[55] Rao, J. N. K. and Molina, I. (2015), *Small Area Estimation, 2nd Edition*, Wiley, ISBN: 978-1-118-73578-7.

[56] Rao, J. N. K. and Molina I. (2016), *Empirical Bayesian and hierarchical Bayes estimation of poverty measures for small areas*, In *Analysis of Poverty Data by Small Area Estimation*, Ed. Pratesi, M., Wiley, 315-324 .

[57] Rubin, D. B. (1981), *The Bayesian bootstrap*, The annals of statistics, 9(1), 130-134.

[58] Sepanski, J. H., and Kong, L. (2007), *A family of generalized beta distributions for income*, arXiv preprint arXiv: 0710.4614.

[59] Tsionas, E. G. (2001), *Exact inference in four-parameter generalized gamma distributions*, Communications in Statistics theory and Methods, 30(4), 747-756.

[60] Toto, M.C.S. and Nandram, B. (2010), *A Bayesian predictive inference for small area means incorporating covariates and sampling weights*, Journal of Statistical Planning and Inference, 140, 11, 2963-2979, Elsevier.

[61] The World Bank (2006), *Nepal - Resilience amidst conflict: an assessment of poverty in Nepal, 1995-96 and 2003-04*, The World Bank.

[62] The World Bank (2013), *Nepal small area estimation of poverty 2011*, vol. 1, Washington DC, The World Bank.

[63] Yu, K., and Zhang, J. (2005), *A three-parameter asymmetric Laplace distribution and its extension*, Communications in Statistics theory and Methods. Series B (Methodological), 34(9-10), 1867-1879.